Journal of Experimental Psychology

1765

Copyright © 1973 by the American Psychological Association, Inc.

DAVID A. GRANT, Editor
University of Wisconsin—Madison

LYLE E. BOURNE, JR., Associate Editor
University of Colorado

HAROLD W. HAKE, Associate Editor
University of Illinois, Urbana

LEO POSTMAN, Associate Editor
University of California, Berkeley

LEONARD E. ROSS, Associate Editor
University of Wisconsin—Madison

DELOS D. WICKENS, Associate Editor
Ohio State University

Consulting Editors

NANCY S. ANDERSON, *University of Maryland*
NORMAN H. ANDERSON, *University of California, San Diego*
MOSHE ANISFELD, *Yeshiva University*
WILLIAM H. BATCHELDER, *University of California, Irvine*
WILLIAM F. BATTIG, *University of Colorado*
ROBERT A. BJORK, *University of Michigan*
GEORGE E. BRIGGS, *Ohio State University*
E. J. CAPALDI, *Purdue University*
VIRGIL R. CARLSON, *National Institute of Mental Health*
HERBERT H. CLARK, *Stanford University*
ANDREW CRIDER, *Williams College*
HELEN B. DALY, *State University of New York College at Oswego*
ROGER L. DOMINOWSKI, *University of Illinois at Chicago Circle*
SHELDON M. EBENHOLTZ, *University of Wisconsin—Madison*
WARD EDWARDS, *University of Michigan*
HOWARD EGETH, *Johns Hopkins University*
BRUCE R. EKSTRAND, *University of Colorado*
HENRY C. ELLIS, *University of New Mexico*
JAMES R. ERICKSON, *Ohio State University*
MURRAY GLANZER, *New York University*
JUDITH GOGGIN, *University of Texas at El Paso*
FRANCES K. GRAHAM, *University of Wisconsin—Madison*
WILLIAM W. GRINGS, *University of Southern California*
DOUGLAS HINTZMAN, *University of Oregon*
JULIAN HOCHBERG, *Columbia University*
LEONARD M. HOROWITZ, *Stanford University*
WILLIAM C. HOWELL, *Rice University*
STEWART H. HULSE, *Johns Hopkins University*
GEOFFREY KEPPEL, *University of California, Berkeley*
GREGORY A. KIMBLE, *University of Colorado*
HERBERT D. KIMMEL, *University of South Florida*
PATRICK R. LAUGHLIN, *University of Illinois*
HERSCHEL W. LEIBOWITZ, *Pennsylvania State University*
FRANK A. LOGAN, *University of New Mexico*
IRVING MALTZMAN, *University of California, Los Angeles*
DONALD R. MEYER, *Ohio State University*
RICHARD B. MILLWARD, *Brown University*
BENNET B. MURDOCK, JR., *University of Toronto*
RAYMOND S. NICKERSON, *Bolt, Beranek, and Newman, Inc., Cambridge, Massachusetts*
MERRILL E. NOBLE, *Pennsylvania State University*
J. BRUCE OVERMIER, *University of Minnesota*
MICHAEL I. POSNER, *University of Oregon*
WILLIAM F. PROKASY, *University of Utah*
L. STARLING REID, *University of Virginia*
DONALD A. RILEY, *University of California, Berkeley*
HERBERT RUBENSTEIN, *Lehigh University*
WILLARD N. RUNQUIST, *University of Alberta*
RUDOLPH SCHULZ, *University of Iowa*
MARIAN SCHWARTZ, *University of Wisconsin—Madison*
BRYAN E. SHEPP, *Brown University*
J. RICHARD SIMON, *University of Iowa*
PAUL SLOVIC, *Oregon Research Institute*
JANET T. SPENCE, *University of Texas at Austin*
HERBERT H. STENSON, *University of Illinois at Chicago Circle*
DAVID R. THOMAS, *University of Colorado*
MILTON A. TRAPOLD, *University of Minnesota*
ENDEL TULVING, *Yale University*
WILLIAM P. WALLACE, *University of Nevada*
WAYNE A. WICKELGREN, *University of Oregon*
GORDON WOOD, *Michigan State University*

HAROLD P. VAN COTT
Managing Editor

ANITA DEVIVO
Executive Editor
and General Manager

HENRY S. SALMON
Advertising and Business
Manager

LOUISE ROSENBLATT
Supervising Technical
Editor

SUSAN BUNKER
DEBORAH FELDER
Technical Editors

ANNE REDMAN
Subscription Manager

Bureau [illegible] Research
6.3.81
Acc... J 710

Volume 98, 1973

$10.10

PUBLISHED BY THE
AMERICAN PSYCHOLOGICAL ASSOCIATION, INC.
1200 SEVENTEENTH ST. N. W., WASHINGTON, D. C. 20036

CONTENTS OF VOLUME 98

Anisfeld, Moshe. *See* Goldfarb, Connie.
Arbuckle, Tannis Y., and Aznavour, Louise. Effectiveness of supplied mediators in relation to presentation modality and retrieval cue 286
Atkinson, Janette. Properties of human visual orientation detectors: A new approach using patterned afterimages 55
Atkinson, Richard C. *See* Wescourt, Keith T.
Aznavour, Louise. *See* Arbuckle, Tannis Y.
Basden, David R. Cued and uncued free recall of unrelated words following interpolated learning 429
Bee, Joan. *See* Kroll, Neal E. A.
Beh, Helen C., and Hawkins, Carole A. Effect of induced muscle tension on acquisition and retention of verbal material 206
Björkstrand, Pär-Åke. *See* Öhman, Arne.
Bourne, L. E., Jr., and Miller, Shirley. Some stimulus variables affecting solution shift performance 291
Bransford, John D. *See* Johnson, Marcia K.
Brogden, W. J. *See* Fingeret, Allan L.
Brooks, David Heber. *See* Nelson, Douglas Lee.
Brown, Ann L. Temporal and contextual cues as discriminative attributes in retardates' recognition memory 1
Brown, Gregory, G. *See* Peterson, Margaret Jean.
Chase, William G. *See* Smith, Edward E.
Cohen, Jeffrey M. Drive level effects on the conditioning of frustration 297
Cook, Cheri L. *See* Hopkins, Ronald H.
Corballis, Michael C., and Miller, Avrum. Scanning and decision processes in recognition memory 379
Cox, Abigail K. *See* Hawkins, Harold L.
Craft, John L. A two-process theory for the short-term retention of motor responses 196
Crowder, Robert G. Representation of speech sounds in precategorical acoustic storage 14
Daly, Helen B., and McCroskery, James H. Acquisition of a bar-press response to escape frustrative nonreward and reduced reward 109
Dawson, Michael E., and Dunn, Fred W. Situational generality of conditioning phenomena: Stimulus generalization and component interactions 440
Dawson, Michael E., and Reardon, Paul. Construct validity of recall and recognition postconditioning measures of awareness 308
Dean, Jeffrey. *See* Ley, Ronald.
Disch, Kenneth. *See* Lappin, Joseph S.
Dunn, Fred W. *See* Dawson, Michael E.
Dyer, Frederick N. Same and different judgments for word–color pairs with "irrelevant" words or colors: Evidence for word-code comparisons 102
Dyer, Frederick N., and Severance, Laurence J. Stroop interference with successive presentations of separate incongruent words and colors 438
Edwards, Richard E. *See* Hopkins, Ronald H.
Edwards, Ward. *See* Fryback, Dennis G.
Eger, Carol E. *See* Peterson, Margaret Jean.
Ellis, Hadyn D. An examination of the adaptation-level theory account of time errors 160
Ellis, Haydn D. Proactive effects of interpolated anchors 233
Ellis, Henry C., and Shumate, E. Chandler. Encoding effects of response belongingness and stimulus meaningfulness on recognition memory of trigram stimuli 70
Ellström, Per-Erik. *See* Öhman, Arne.
Ericcson, Stig. *See* Lisper, Hans-Olof.
Esposito, Nicholas J. Effects of stimulus novelty and dimensional saliency in human shift learning 264
Fingeret, Allan L., and Brogden, W. J. Effect of pattern display by letters and numerals upon acquisition of serial lists of numbers 339
Finkelman, David G. *See* Spector, Amos.
Forbach, Gary B. *See* Stanners, Robert F.
Fryback, Dennis G., Goodman, Barbara C., and Edwards, Ward. Choices among bets by Las Vegas gamblers: Absolute and contextual effects 271
Gambino, Blase. *See* Schwartz, Barry J.
Gatchel, Robert J., and Lang, Peter J. Accuracy of psychophysical judgments and physiological response amplitude. 175
Gildow, Joseph W. *See* Melamed, Lawrence E.
Goldfarb, Connie, Wirtz, Joyce, and Anisfeld, Moshe. Abstract and concrete phrases in false recognition 25
Gollob, Harry F., and Lugg, Andrew M. Effects of instruction and stimulus presentation on the occurrence of averaging responses in impression formation 217
Goodman, Barbara C. *See* Fryback, Dennis G.
Goodrick, Charles L. Maze learning of mature-young and aged rats as a function of distribution of practice 344
Gould, John D. Eye movements during visual search and memory search 184
Gursky, Gabriele. *See* Kroll, Neal E. A.
Halay, Michael. *See* Melamed, Lawrence E.
Hawkins, Carole A. *See* Beh, Helen C.
Hawkins, Harold L., McDonald, Gary J., and Cox, Abigail K. Effects of irrelevant information in speeded discrimination 435
Hodge, Milton H., and Pennington, Florrie M. Some studies of word abbreviation behavior 350
Hopkins, Ronald H., Edwards, Richard E., and Cook, Cheri L. Presentation modality, distractor modality, and proactive interference in short-term memory 362

CONTENTS OF VOLUME 98

Howell, William C. Storage of events and event frequencies: A comparison of two paradigms in memory 260
Hung, Daisy L. *See* Tzeng, Ovid J. L.
Jarvella, Robert J. Coreference and short-term memory for discourse 426
Johnson, Marcia K., Bransford, John D., and Solomon, Susan K. Memory for tacit implications of sentences 203
Jones, Mari Riess. *See* Shinar, David.
Katz, Stuart. Role of instructions in abstraction of linguistic ideas 79
Kausler, Donald H. Homographs as double function items in verbal discrimination learning 214
Kelly, Richard T. *See* Martin, David W.
Kennedy, Alan. Decision latencies to thematic and nonthematic distractors in prose 432
Kroll, Neal E. A., Bee, Joan, and Gurski, Gabriele. Release of proactive interference as a result of changing presentation modality 131
Krueger, Lester E. Effect of irrelevant surrounding material on speed of *same–different* judgment of two adjacent letters 252
Lang, Peter J. *See* Gatchel, Robert J.
Lappin, Joseph S., and Disch, Kenneth. Latency operating characteristic: III. Temporal uncertainty effects 279
Laughery, Kenneth R. *See* Spector, Amos.
Laughlin, Patrick R. Focusing strategy in concept attainment as a function of instructions and task complexity 320
Levison, Moshe J., and Restle, Frank. Effects of blank-trial probes on concept-identification problems with redundant relevant cue solutions 368
Levy, Robert M. Redundancy and dimensional load in stimulus identification 138
Ley, Ronald, and Dean, Jeffrey. Effects of associative reaction time and meaningfulness in free recall of mixed and unmixed lists 220
Lieury, Alain. Facilitation, inhibition, and distortions of the serial curve in single-trial free recall as a function of prior within-word organization 91
Lisper, Hans-Olof, and Ericsson, Stig. Effects of signal frequency on increase in reaction time in a 10-minute auditory monitoring task 316
Lively, Barry L. *See* Schuberth, Richard E.
Lordahl, Daniel S. *See* Schwartz, Barry J.
Lugg, Andrew M. *See* Gollob, Harry F.
Marston, Paul T. *See* Martin, David W.
Martin, David W., Marston, Paul T., and Kelly, Richard T. Measurement of organizational processes within memory stages 387
Massaro, Dominic W. *See* Thompson, Muriel C.
McConkie, Alan. *See* Shulman, Harvey G.
McCroskery, James H. *See* Daly, Helen B.
McDonald, Gary J. *See* Hawkins, Harold L.
McIntyre, John S., Stojak, R. A., and Mostoway, W. Individual organization and release from proactive interference 164
Melamed, Lawrence E., Halay, Michael, and Gildow, Joseph W. Effect of external target presence on visual adaptation with active and passive movement 125
Miller, Avrum. *See* Corballis, Michael C.
Miller, Shirley. *See* Bourne, L. E., Jr.
Mostoway, W. *See* McIntyre, John S.
Nelson, Douglas Lee, and Brooks, David Heber. Functional independence of pictures and their verbal memory codes 44
Nickerson, R. S., and Pew, R. W. Visual pattern matching: An investigation of some effects of decision task, auditory codability, and spatial correspondence 36
Northman, John E., and Ranken, Howard. Selective attention in the effect of labeling on symbolic problem solving 209
Öhman, Arne, Björkstrand, Pär-Åke, and Ellström, Per-Erik. Effect of explicit trial-by-trial information about shock probability in long interstimulus interval GSR conditioning 145
Pennington, Florrie M. *See* Hodge, Milton H.
Peterson, Margaret Jean, Eger, Carol E., and Brown, Gregory G. Meaningfulness, phonemic similarity, and sensory memory 64
Pew, R. W. *See* Nickerson, R. S.
Ranken, Howard. *See* Northman, John E.
Reardon, Paul. *See* Dawson, Michael E.
Restle, Frank. *See* Levison, Moshe J.
Reutener, Donald B. *See* Schuberth, Richard E.
Richardson, Jack. Effect of speed of learning and degree of learning on cue selection 396
Runquist, Peggy A. Transfer of response groupings in paired-associate learning 152
Runquist, Willard N. Differences in coding processes responsible for interference in paired-associate learning 404
Saba, Anton K., and Turnage, Thomas W. Unit-sequence interference in short-term memory: Facilitation versus interference factors 328
Sanders, Nicholas M., and Tzeng, Ovid J. L. Learning outcomes and the learner's consistency seeking in rote and conceptual learning 302
Schuberth, Richard E., Lively, Barry L., and Reutener, Donald B. Release from proactive interference in the recall of sentences 423
Schwartz, Barry J., Lordahl, Daniel S., and Gambino, Blase. Effect of sequence structure on recall 212
Sekuler, Robert. *See* Steinberg, William.
Severance, Laurence J. *See* Dyer, Frederick N.
Shinar, David, and Jones, Mari Riess. Effects of set-inducing instructions on recall from dichotic inputs 239
Shulman, Harvey G., and McConkie, Alan. S–R compatibility, response discriminability, and response codes in choice reaction time 375

Shumate, E. Chandler. *See* Ellis, Henry C.

Smith, Edward E., Chase, William G., and Smith, Peter G. Stimulus and response repetition effects in retrieval from short-term memory: Trace decay and memory search 413

Smith, Peter G. *See* Smith, Edward E.

Solomon, Susan K. *See* Johnson, Marcia K.

Spector, Amos, Laughery, Kenneth R., and Finkelman, David G. Rehearsal and organization in intentional forgetting 169

Stanners, Robert F., and Forbach, Gary B. Analysis of letter strings in word recognition 31

Steinberg, William, and Sekuler, Robert. Changes in visual spatial organization: Response frequency equalization versus adaptation level 246

Stojak, R. A. *See* McIntyre, John S.

Thompson, Muriel C., and Massaro, Dominic W. Visual information and redundancy in reading . 49

Townsend, V. M. Loss of spatial and identity information following a tachistoscopic exposure .. 113

Turnage, Thomas W. *See* Saba, Anton K.

Tzeng, Ovid J. L. *See* Sanders, Nicholas M.

Tzeng, Ovid J. L., and Hung, Daisy L. Intralist organization and subsequent free recalls 119

Uhlarik, John J. Role of cognitive factors on adaptation to prismatic displacement 223

Vroon, A. G. *See* Vroon, P. A.

Vroon, P. A., and Vroon, A. G. Tapping rate and expectancy in simple reaction time tasks 85

Wescourt, Keith T., and Atkinson, Richard C. Scanning for information in long- and short-term memory 95

Wickens, Thomas D., and Zax, Brian. Asymmetric attribute interaction in concept identification . 335

Wirtz, Joyce. *See* Goldfarb, Connie.

Zax, Brian, *See* Wickens, Thomas D.

Not just...
"another pretty face."

When you look at the new BRS/LVE electromechanical control system, you're looking at the oldest new line in the field. Its predecessor, the old Lehigh Valley Electronics and BRS-Foringer line was used as the model. We took the best features of both lines, combined and refined them even further, used the best and most reliable components, improved the operating specs and came up with a line that is second to none—Anywhere.

As one example, we retained the rugged lamacoid used to make-up the panel faces. Why lamacoid? Well, for the best reasons of all—It's extremely strong and it resists chipping, peeling and scratching better than any other substance we've ever tried. It costs more for us to use but it will hold up even under the worst of conditions.

We also knew we had to make sure the new line was **100 percent compatible** with all of the existing BRS-Foringer and Lehigh Valley Electronics relay equipment still being used in labs throughout the world.

It is.

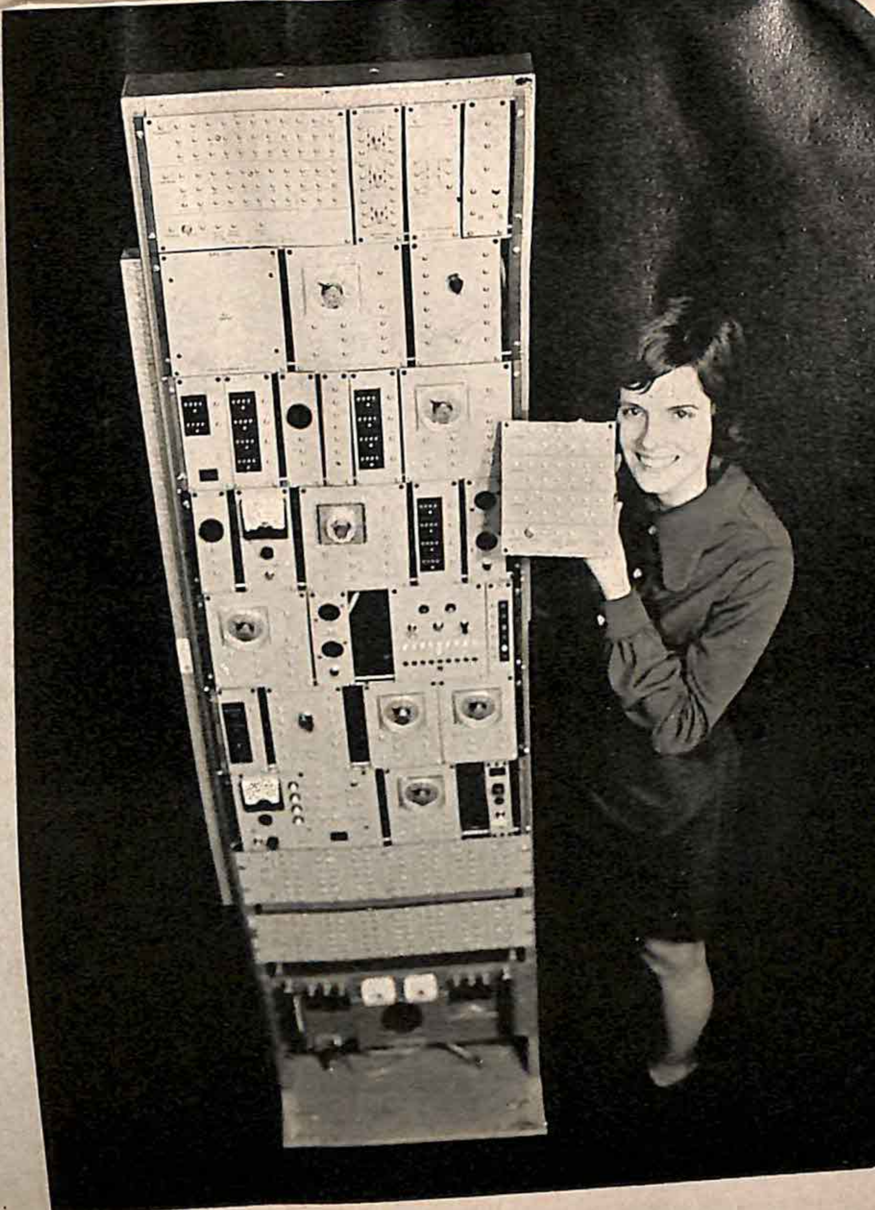
Further, we had to develop the most complete line of support equipment to be found in the field.

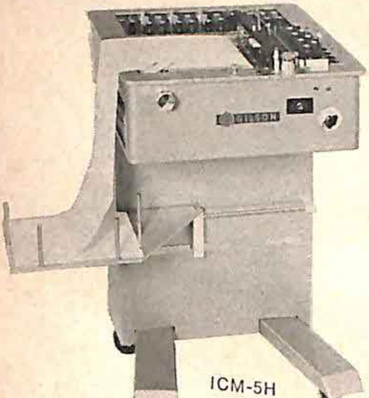
It is.

So, when you use the new BRS/LVE electromechanical control system, you're using equipment with not just "ANOTHER PRETTY FACE". It has more than 40 years of combined company know-how behind it.

For a complete listing of our entire electromechanical line, please write for our latest catalogs.

BRS/LVE
DIVISION of TECH SERV inc.
5301 Holland Drive • Beltsville, Md. 20705
Telephone 301/474-3400





ICM-5H

GILSON BISCRIPTUAL POLYGRAPHS

NEW! HEATED STYLUS PHYSIOLOGICAL RECORDING ON THERMOCHROMIC PAPER

Bond paper texture • Abrasion-resistant • Crumple-resistant (for the record that didn't look relevant) • Substantially lower in cost than conventional wax-coated heat-sensitive paper

NEW! HIGH FREQUENCY PEN RESPONSE UP TO 90 Hz

NEW! INTEGRATED CIRCUIT ELECTRONICS

5- and 8-channel Polygraphs
5-channel Projector Polygraph
1- and 2-channel portable recorders

50 mm deflection galvanometer channels
Pressure, force, direct coupled potentials, respiration, ECG, EEG, EMG, etc.

310 or 200 mm: full chart width servo channels
Dye dilution curves, GSR, O_2, pH, CO_2, temperature, respiration, etc.

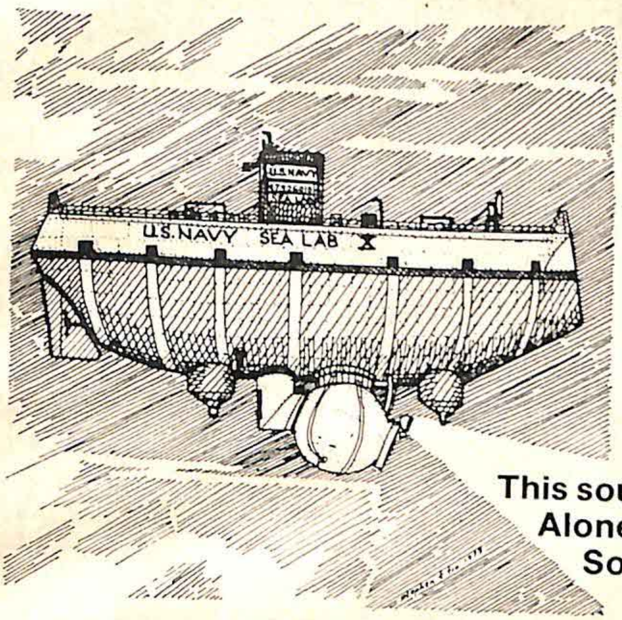
Biscriptual Polygraphs can also be converted for ink pen recording

Call or write for literature and a sample recording

GILSON®

GILSON MEDICAL ELECTRONICS, INC.
P.O. BOX 27, MIDDLETON, WISCONSIN 53562 • TELEPHONE 608/836-1551
EUROPEAN MANUFACTURING PLANT
Gilson Medical Electronics • 69, rue Gambetta, 95-Villiers-le-Bel, FRANCE • Telephone 990-10-38

Aldine Books for Spring


U.S. NAVY SEA LAB X
U.S. NAVY

This soul hath been
Alone on a wide wide sea;
So lonely 'twas, that God himself
Scarce seemed there to be.
Coleridge

MAN IN ISOLATION AND CONFINEMENT

John E. Rasmussen, *editor*

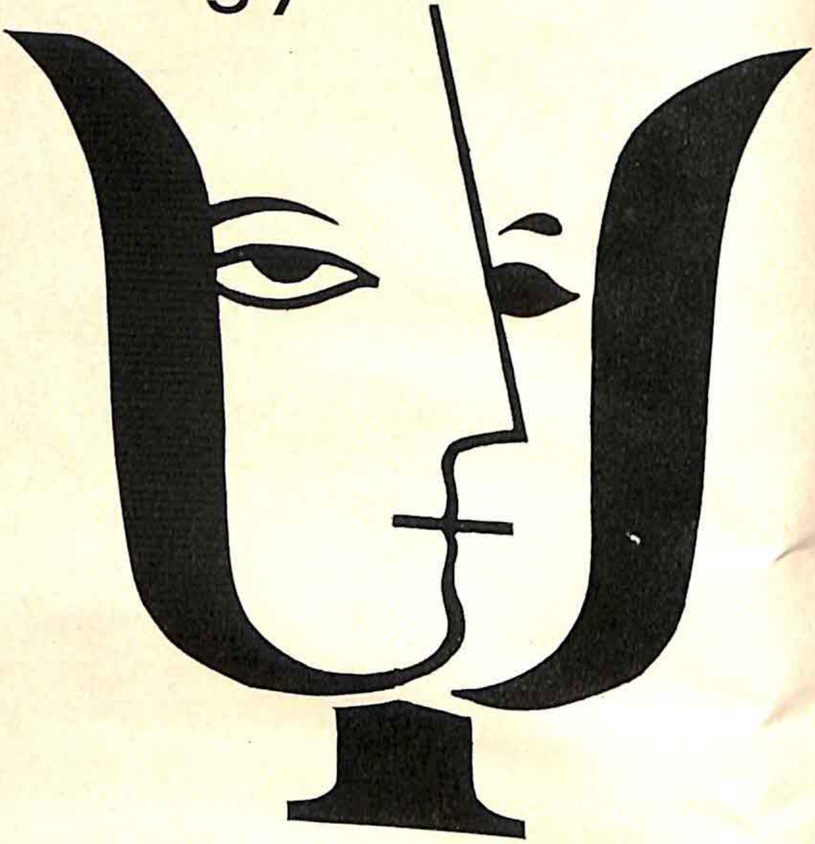
This book focuses particularly on those special circumstances in which men (alone or in groups) are isolated or confined for periods of time long enough to affect the way in which they think and behave. The first attempt to cover the total spectrum of isolation and confinement in one volume, the book is based upon a substantial expansion and revision of the papers presented in 1969 at the Rome Symposium of Man in Isolation and/or Enclosed Space, sponsored by the NATO Science Committee Advisory Group on Human Factors.

The chapters are arranged so as to begin with study of the individual, proceed through artificial and natural groups, and conclude with broad ecological and taxonomic considerations. Highlights include: A summary and critical evaluation of work in the field of sensory deprivation, a report of the classic French cave studies, a summary of studies of personality development of natural isolated groups in northern Norway, two syntheses of fifteen years of research dealing with behavior in Antarctic isolation, two chapters venturing bold theoretical formulations on the basis of highly controlled laboratory studies of isolation and confinement at the U.S. Navy Medical Research Institute and an essay on the questions of taxonomy that arise from the study of these phenomena.

320 pp. $9.50 cloth June, 1973

Aldine Publishing Co.
529 South Wabash Avenue
Chicago, Illinois 60605

professional
psychology



Professional Psychology is devoted to the application of principles and skills in a wide variety of settings: clinics, hospitals, schools, industry, government, and independent practice. The journal contains original articles on both theory and techniques which will enhance knowledge and improve the effectiveness of psychologists. Quarterly: one volume per year. The 1973 volume is Vol. 4. Subscriptions: $12.00 (foreign $13.00).

American Psychological Association
1200 Seventeenth Street, N.W.
Washington, D.C. 20036

• Payment must accompany orders amounting to $15.00 or less •

Diary No.....718
Date 328
Bureau Ed & Pay Resea

Journal of
Experimental Psychology

Copyright © 1973 by the American Psychological Association, Inc.

VOL. 98, NO. 1 APRIL 1973

TEMPORAL AND CONTEXTUAL CUES AS DISCRIMINATIVE ATTRIBUTES IN RETARDATES' RECOGNITION MEMORY[1]

ANN L. BROWN[2]

Children's Research Center, University of Illinois

Moderately retarded adolescents served as *S*s in a series of five experiments investigating the cues involved in recognition memory for pictures. The experiments were designed to compare strength, frequency attribute, and context marker theories of recognition memory. Accuracy at distinguishing target (T) items from distractor (D) items exceeded 90% correct even when all items were equated for frequency of occurrence and recency of last appearance. As long as the first occurrence of the items was temporally isolated in an inspection set this level was maintained. However, when both T and D items occurred initially together in the inspection set, performance dropped to 70% correct. The data were seen as inconsistent with strength and frequency attribute theories and consistent with a context marker theory of recognition. The special role of temporal cues was discussed, with particular emphasis on the importance of the item's "birthdate," the time tag associated with the item's initial appearance.

It has recently been established that young children and retardates perform extremely efficiently on recognition memory tasks (Brown, 1972; Brown & Campione, 1972; Brown & Scott, 1971). The principal aim of this research is to investigate the processes involved in retardates' recognition memory for pictures. Specifically, this series of studies was designed to compare a strength theory (Kintsch, 1970) with a marker theory (Anderson & Bower, 1972) of recognition memory. Two marker theories were considered: Anderson and Bower's (1972) context marker theory and Underwood and Freund's (1970) frequency attribute theory. The dominant strategy throughout the series of experiments was to engineer situations where recognition performance could not be based on a simple strength comparison in an attempt to discover what other attributes could be used to differentiate items in memory. The main attributes investigated were frequency, context, and time of occurrence.

Consider first a simple recognition memory task. Items are presented for study and are then mixed with new items. The *S* is required to identify the study items. Performance can be based upon a frequency lifferential, the old items having a frequency of one, the new items of zero. Vhile it is clear that frequency is an important factor, the concept of situational frequency (Underwood & Freund, 1970) must be introduced in an attempt to explain recognition performance. For it is equally clear that in the standard recognition task, the stimuli used have been seen or heard before the experimental list is presented. Pictures cut from popular magazines and children's books have typically been used in studying recognition memory in children. These items are all in some sense familiar to *S*s, who spontaneously indicate that they have seen an item before ("at home in my Snow White book") but still choose the familiar item as new within the experimental setting (Brown & Scott, 1971). Furthermore, retardates are able to distinguish between items seen within the experimental setting but at different times (Brown, 1972). In order to do this, *S*s must have some means of determining not only that an item has been seen before, but where, when, or how it occurred. As Anderson and Bower (1972) have recently pointed out, what appears to be a simple recognition memory experiment is actually "a list-differentiation experiment in disguise [p. 102]."

---

[1] This series of studies was conducted at the University of Connecticut and supported by U.S. Public Health Service Grants HD-06864 and HD-05951 and by Research Grant M-1099 from the National Institute of Mental Health. The research was carried out at the Mansfield State Training School, Mansfield Depot, Connecticut. The author expresses appreciation for the active cooperation of Superintendent Francis P. Kelly and Director of Training Louis Boly. The author would also like to express her appreciation to David Zeaman for his advice and support throughout the conduct of the research and Richard Sperber for his help and advice in collecting the data for the first two experiments.

[2] Requests for reprints should be sent to Ann L. Brown, Children's Research Center, University of Illinois, Champaign, Illinois 61820.

The question remains: What attributes of memory does *S* use to enable him to make the requisite discrimination of old and new. The dominant theory of recognition memory has been the strength theory, reviewed in detail by Kintsch (1970). For the purposes of the present article it is sufficient to state briefly that the strength theory assumes that *S* can directly access a memory trace of a previously presented item and make recognition decisions about that item "not on the basis of a detailed memory of the past history of occurrences of the item, but rather on the basis of a single measure which reflects to some extent its past frequency, recency, and duration of exposure [Anderson & Bower, 1972, p. 98]." Thus the essential feature of the strength theory of recognition memory is that it is "ahistorical." Therefore, it should be impossible to distinguish between items which have been equated for frequency and recency and, thus, have the same resultant strength in memory, but have in some other respects different past histories.

An alternative to a strength theory of recognition memory is a context marker theory such as that recently proposed by Anderson and Bower (1972). This theory assumes that items in memory do have histories and that, in fact, items are recognized precisely by retrieving certain information concerning their histories, such as contextual information which is stored together with the item (Tulving & Thomson, 1971). The essential feature of this theory is, therefore, the antithesis of strength theories in that it assumes that the historical features of a memory trace are maintained and can be used to differentiate items in memory.

The third theory which will be considered is also a marker or attribute theory. Underwood and Freund (1970) assume that a frequency differential is the dominant memory attribute in recognition memory tasks. A simple frequency theory could be viewed as a strength theory, since increasing an item's frequency of occurrence is one of the main methods of incrementing its memory strength (Kintsch, 1970). However, Underwood and Freund are not traditional strength theorists, as they assume that frequency is only one of many memory attributes and that other attributes (such as context) can come to support a frequency discrimination or even replace it when frequency as a basis for performance becomes invalid. The theory is distinct from general marker theories in that it assumes a hierarchy of attributes with frequency dominant. If frequency is removed, performance should decline as other, less dominant cues must be used.

The three theories result in contrary predictions. For example, strength theories (Kintsch, 1970) would assume that accuracy could not be maintained unless the items differed in terms of some variable known to influence their strength (e.g., frequency or recency). Similarly, if frequency were the dominant cue, equating items for frequency should make it more difficult for *S*s to discriminate betwee items. However, context marker theories (Anderson & Bower, 1972) would assume that as long as the items to be differentiated carry distinct contextual markers *S* will be able to discriminate items in memory even in the absence of frequency information.

EXPERIMENT I

In order to distinguish between the opposing views presented above, it would be necessary to modify a recognition memory task in such a way that item differentiation on the basis of frequency would be impossible. Underwood and Freund (1970), using adult *S*s, compared recognition memory for study items when both the study items and the distractors had occurred equally frequently. Performance should be at chance under these conditions if frequency alone is the basis of performance. However, although performance did decline (between 15%–20%), *S*s were still able to discriminate study items at an above chance level (approximately 70% correct). Martin (1970) has shown that when both study items and distractors were repeated many times, retardates were still able to discriminate study items which had occurred five times from distractors which had occurred four times. However, no discriminations between items of equal frequency were included in Martin's study.

In the present study, the Underwood and Freund (1970) procedure of pairing study nd distractor items equated for experiental frequency was adopted. In addiion, the other feature which contributes o an item's strength in memory, its exerimental recency, was also controlled in n attempt to determine whether items in emory can be differentiated on the basis f some other cue than their experimentally efined strength, or their frequency of ccurrence.

Method

ubjects

The *S*s were 18 institutionalized, moderately rerded *S*s selected from the Mansfield State Traing School, Mansfield Depot, Connecticut, without gard to sex or diagnostic category. Six *S*s failed etraining and two *S*s were not available for the tire experiment so their data were discarded. The remaining 10 *S*s ranged in age 12 yr. 3 mo.–18 yr. 5 mo. with a mean chronological age (CA) of 16 yr. 5 mo. Their mental age (MA) range was 8 yr. 2 mo.–12 yr. 4 mo. with mean MA of 9 yr. 9 mo. They had a mean IQ of 71, with a range of 58–83.

Apparatus

The stimuli were presented by means of a modified Wisconsin General Testing Apparatus (WGTA). The *S* and *E* sat facing each other with a one-way vision screen interposed. The main features of the apparatus consisted of a 76.2 × 30.5 cm. sliding tray with two (6.4-cm.) reward wells centered 30.5 cm. apart. The sliding tray was pulled out of sight behind the one-way vision screen for baiting and pushed out directly in front of *S* for each trial.

Stimulus Materials

A large stimulus pool was constructed consisting of colorful pictures cut from children's books and magazines. The pictures were pasted onto 10.2 × 10.2 cm. black posterboard cards and covered with transparent Contact paper. With the exception of experimentally determined repetitions, no picture was repeated and an entirely new set of pictures was used each day. During pretraining the pictures were selected randomly, without replacement, from a separate large pool of items. During the experimental sessions, five preordered sets were available, one for each day.

Before describing the design and procedure a note is in order concerning the terminology to be used in these experiments. The to-be-remembered items will be referred to as target items (T items). The new items on forced-choice tests will be referred to as distractor items (D items). On its first occurrence an item will be given the subscript zero. Each subsequent appearance will increase the subscript by one. Therefore, D_1 will indicate a distractor item that has occurred once before and D_2 will indicate twice before, etc. Thus, a new item will always have a zero subscript to indicate that it has never occurred before and an old item will always have a nonzero subscript, the number referring to its number of previous appearances.

Design

Frequency controlled trials. Each experimental sequence consisted of an inspection series of 36 items followed by a test series of 52 forced-choice pairs. Frequency was varied in three blocks of 12 trials. On Trials 1–12, each pair of items contained a T item selected randomly from the 36 available T items of the inspection set. As it had occurred before in the inspection series it was designated T_1. Each T_1 item was paired with a distractor item that had never occurred before, D_0. Thus, the T_1–D_0 discrimination on Trials 1–12 was a standard recognition memory test with an old item paired with a new item. On Trials 13–24 the forced-choice pairs again consisted of T_1 items, randomly selected from

TABLE 1

Design and Results of Experiments I and III

Trials	Problem type[a]	No. of observations		Proportion correct	
		Per day	Total	Experiment I	Experiment III
Test trials					
1–12	T_1–D_0	12	600	.95	.97
13–24	T_1–D_0	4	200	.92	.92
	T_1–D_1	8	400	.87	.90
27–38	T_1–D_0	4	200	.90	.94
	T_1–D_1	2	100	.91	.90
	T_1–D_2	6	300	.91	.93
25–26	T_2–D_2	2	100	.95	—
39–40	T_2–D_2	2	100	.94	—
Recognition of D items (Experiment I only)					
41–44	T_3–D_0	4	200	.99	—
45–52	D_1–D_0	4	200	.95	—
	D_3–D_0	4	200	.96	—

Note. Prior to the test trials, an inspection set of 36 items (in a stack in Experiment I and in pairs in Experiment III) was presented to *S*s.

[a] T = target item; D = distractor item; and subscript gives number of previous appearances of item.

the remaining inspection set pool. These were paired randomly with one of two types of D items, those that had occurred as distractors on Trials 1–12 or new D items. Thus, in a resultant T_1–D_1 pair, both items had occurred before, equally often. Discrimination based on frequency alone would be impossible. On Trials 27–38, the T_1 items, consisting of the remaining 12 items not yet tested from the inspection set, were paired with one of three types of distractors. The D item was either entirely novel, D_0; had occurred once before, D_1; or twice before, D_2, as a distractor. Thus, in the T_1–D_2 condition, both items were old, but now the D item had occurred more frequently. If *S* chose the one occurring more frequently, he would choose the D item. In the third block of trials the three frequencies of D items occurred randomly. The sequence of experimentally controlled frequency pairs is presented in Table 1.

Recency controlled trials. The remaining trials in the first sequence of 40 were included as a test of whether *S* used relative recency of occurrence in order to discriminate between T and D items. On Trials 25 and 26, 39 and 40, a T_2–D_2 pair was presented, both items occurring for the third time. These items were selected in such a way that T_1 and D_1 items from approximately the same position in the preceding block of 12 trials were repeated now as T_2–D_2 pairs. The resultant T_2–D_2 pair would provide a choice between two items, both of which had occurred equally often, with their last appearance being at approximately the same time. If *S* based his decision on either frequency, or recency of last appearance of an item, performance would be at chance.

Recognition of distractor items. The remaining trials of the experimental sequence were included as a control condition, the purpose of which was to test whether *S*s were able to recognize a D item if it was paired with an entirely new picture. For, although the T and D items had occurred an equal number of times in the frequency controlled pairs, they may not have been considered equally if *S* failed to attend to the D items. Consider the trials on which *S* looked first at the T item. If he identified this as the old item he may not have considered the D item closely before making his choice. Thus, an experimentally defined frequency of occurrence would not correspond to a subjective frequency count. In order to determine whether the D items were encoded at all, it is necessary to show that they can be distinguished from totally new items. Therefore, on Trials 41–52 the instructions to *S* were altered. He was told that the game had changed and that now he must choose the new picture: "The candy will be under the one you never saw before." In order to ensure that *S*s understood the changed instructions they were presented on four trials (41–44) with a choice between a very familiar T item (T_3) and a totally new distractor (D_0). On the remaining trials (45–52) *S*s were presented with two D items, a completely novel D item (D_0) paired with a D item that had occurred either once before (D_1) or three times before (D_3) as a distractor item. There were four tests of each comparison (see Table 1).

In order to arrive at the necessary sequence of item repetition and to ensure that item type was not completely confounded with position in the test sequence, an unequal number of observations of each trial type occurred within a day. In the first section (Trials 1–40) there were 20 T_1–D_0 tests, 10 T_1–D_1 tests, and 6 T_1–D_2 tests. With 10 *S*s, tested for 5 days each, there were totals of 1,000, 500, and 300 observations of the three problem types. The order of presentation, together with the number of observations per day, is presented in Table 1.

Procedure

Each *S* was tested for a total of 8 days, 3 days of pretraining followed by 5 experimental sessions.

Pretraining. The purpose of the pretraining sessions was to train *S*s to choose the T or D items on command. On the initial day of pretraining *S* were trained on a forced-choice recognition memory procedure. They were given an inspection set o items and told that these were "candy cards." Following this, *S*s were given a series of forced-choice tests where a choice between a candy car (T item) and a novel distractor (D item) was presented. The *S*s were told to choose the candy card On the initial day *S*s received 10 trials with an inspection set of only one item, followed by one tria with an inspection set of 2, 5, and finally 10 item On the second day the procedure was repeated wit the exception that *S*s were instructed on the forced-choice tests to choose the D item: "The candy wil be under the one you never saw before." On the third day *S*s were required to choose the T or D item randomly on command. On all forced-choice tests the correct item was placed over the reward well which was baited with an M&M candy. The position of the correct item was determined by a Gellermann (1933) series. Success criterion was set at 100% correct, but *S*s were allowed one repetition under any condition in which they made an error.

Experimental session. The 5 days of experimental sessions were identical, with the exception that *S*s received a new set of preordered cards each day. On each day *S*s were given an inspection set of 36 items and told that these were the candy cards. The *S* and *E* sat together in front of the WGTA and looked at the pictures. The *S* looked at the items one at a time at his own rate, but approximately 3 sec. were required to view each picture. Following the viewing session, *E* then went behind the WGTA screen and began the forced-choice trials. The *S* was instructed to find the candy card. The sequence of 52 forced-choice tests followed in the prearranged order, with the position of the correct card determined by a Gellermann (1933) sequence. Throughout the series of trials, correct choices were rewarded with an M&M candy. At the end of each day's testing each *S* received an additional candy reward of his choice.

Results and Discussion

The results for all test types are shown in Table 1.

Frequency Controlled Trials

Performance was high on the initial set of 12 test pairs (Trials 1–12), where a standard recognition memory task was involved. The probability of correctly identifying the old picture was .95. A similar accuracy level (.96) has previously been reported in both a continuous recognition memory paradigm (Brown, 1972) and a forced-choice procedure (Martin, 1970) with this *S* population.

On the remaining trials, where D items begin to recur, performance dropped to approximately 90% correct. The block of trials most suitable for comparison is the third block (27–38), for here all types of T–D comparisons occur randomly. Performance levels are equal for all comparisons. These data are inconsistent with those reported by Underwood and Freund (1970), where a significant decline in performance was found when frequency equated pairs were presented.

Recency Controlled Trials

Performance on the test trials (Trials 25, 26, 39, and 40), where both frequency and recency of occurrence were equated, was again very accurate (94%). As the items were approximately equated for strength in that neither differential frequency nor recency could serve as a basis for discrimination, the absence of a decline in accuracy is difficult to explain in terms of a standard strength theory of recognition memory. Also difficult for a strength theory to explain is that performance was also very accurate when T_1–D_2 pairs were presented. The D items were seen more frequently and should therefore have greater strength. In order for a strength theory to incorporate these data, complex decision rules concerning when to approach and when to avoid the strong/weak item would have to be added.

Recognition of Distractor Items

Accuracy at distinguishing recurring D items from entirely novel D items was high (96% correct). This suggests that *S* must encode the D items, if he can distinguish between an item that occurred once as a distractor (as many as 40 trials previously) from a new item.

Considered as a whole, the data suggest that neither a traditional strength model of recognition memory nor a frequency attribute model can fully account for performance. The possibility that contextual or temporal features are used to differentiate items will be considered later. However, before this was attempted, one additional explanation of the present data was considered. That is, in this paradigm *S* could learn, via a contiguity mechanism, which items are correct, with this serving as the basis of his accurate performance. This possibility was examined in Experiment II.

EXPERIMENT II

The data from Experiment I are at variance with those reported by Underwood and Freund (1970) in that there was no decline in performance associated with presenting items of equal frequency. One possible reason is that the Underwood and Freund study used a large number of test pairs (200). As a consequence, the frequency controlled pairs were embedded in a long series of standard recognition pairs. The *S*s could not easily learn that distractor words were being repeated. This excellent control was not possible in Experiment I because of the restrictions placed on the design by the particular *S* population. With retarded *S*s it appeared that a test sequence of 50 was an absolute maximum. Thus, it is obvious that *S*s were not only aware that D items repeated, but it is possible that they could learn which items to avoid within the course of the experimental sequence.

That learning could explain performance was proposed by Martin (1970), who suggested that *S* comes to distinguish T and D items by means of contiguity learning. In Martin's study the same T and D items were constantly re-paired prior to presentation on successive trials. On the initial trial, *S* could choose the T item because the salient novelty dimension was still present. On this trial, *S* would choose the T item because of its familiarity; however, due to this choice, an association would be formed between that T item and an approach response via a contiguity mechanism. These associations were then assumed sufficient to maintain performance when the novelty cue faded. However, in Experiment I of this series, the procedure was such that a new T item occurred on each test trial. Any associations between the T items and an approach response could not help subsequent performance, since the T items were always replaced. An explanation in terms of contiguity would have to be based upon avoidance responses being built up to the D items. Further, these avoidance responses would have to be established in one trial and maintained when the original transverse compounds were destroyed. While this account is possibly correct, it appears unlikely that such is the case. For example, there are data which indicate that not only are avoidance responses formed slowly by retardates but that destroying transverse compounds also has a deleterious effect on retardates' performance (Campione, 1969). Thus, a contiguity explanation appeared unlikely to explain the present data. However, in view of the divergence between the Underwood and Freund (1970) data and that obtained in Experiment I, a further test was made in Experiment II before such a hypothesis could definitely be ruled out. In Experiment II, *S* was required to discriminate between T and D items when both frequency of occurrence and number of rewarded choices were equated.

Method

Subjects, Apparatus, and Stimulus Materials

These were the same as in Experiment I with the exception that all new pictures were used.

Procedure

No additional pretraining was given. The *S*s were tested for a total of 3 days. On each day they received two repetitions of the experimental sequence, and therefore, two separate sets of pictures. At the beginning of each experimental sequence *S* was shown an inspection series of 20 cards and told that these were the candy cards. He viewed the stack at his own rate, one picture at a time. The viewing period was followed by 30 forced-choice test trials. On Trials 1–10 the standard recognition memory procedure was followed by presenting T_1 items, selected randomly from the inspection series, with novel distractors, D_0. The *S*s were told to find the candy card. On the second block of trials (11–20), the remaining 10 T items were selected from the inspection set and paired with the D items that had occurred in the preceding block of trials, yielding T_1–D_1 pairs. However, in this block *S*s were told to choose the card that had not occurred in the inspection series (i.e., the D item). The *S*s had previous experience with picking the T and D items on demand in pretraining. On the last block of 10 trials, the T_1 items that had occurred on Trials 1–10 were randomly re-paired with the 10 D items which had occurred on Trials 11–20, yielding a set of T_2–D_2 pairs. In this block, both the T and the D items had been selected as correct, and rewarded, once before. The T_2 items were correct on Trials 1–10 and the D_2 items were correct on Trials 11–20. Therefore, both the T and the D items had occurred equally often and had been associated with approach responses and rewarded equally often. The sequence of test trials is presented in Table 2.

There were, therefore, three problem types, T_1–D_0, T_1–D_1, and T_2–D_2, and 10 repetitions of each type within an experimental sequence. With two repetitions of a sequence each day and 3 days of testing 10 *S*s, the design yielded 600 observations of each problem type.

Results and Discussion

The results are shown in Table 2. Performance declined only slightly as a function of trial block. The overall probability of a correct response was .95. Even when both frequency of occurrence and approach responses were controlled in the third block of trials, performance remained at 95% accurate.

There was no apparent effect of repetitions so the data were summed across this variable. A 3 × 10 (Trial Blocks × *S*s) factorial analysis of variance was conducted on the error scores. The effect due to trial block was significant, F (2, 18) = 4.17, $p < .05$. Post hoc testing revealed that performance in the first block of 10 trials, the standard recognition test, was significantly better than in the second block of trials, S^2 (2, 18) = 8.22, $p < .05$, but no other pair-wise comparison reached significance. Of principal interest is the finding that performance in the third block of trials is not significantly different from that in the first. Even when both frequency of occurrence and reward history were controlled, performance remained at 95% accurate. These data, together with the results of Experiment I, suggest that contiguity learning is an unlikely mechanism with which to explain the excellent performance.

TABLE 2

Design and Results (Mean Proportion Correct as a Function of Trial Block) of Experiment II

Trials	Problem type[a]		Proportion correct
1–10	$T_1\overset{+}{(1\text{–}10)}$	$D_0\overset{-}{(1\text{–}10)}$	.98
11–20	$T_1\overset{-}{(11\text{–}20)}$	$D_1\overset{+}{(1\text{–}10)}$	.93
21–30	$T_2\overset{+}{(1\text{–}10)}$	$D_2\overset{-}{(1\text{–}10)}$	.95

Note. All proportions are based on 600 observations.

[a] T = target item; D = distractor item; subscript gives number of previous appearances of item; (+) = *S* was told to find the candy card (card he had seen before); (−) = *S* was told to choose the card that he had not seen before; numbers in parentheses give trials on which item was seen before.

Experiment III

Considered together, the data from the first two experiments suggest that neither a strength theory nor a frequency attribute theory can readily explain how the T and D items are differentiated. The third theory to be considered is Anderson and Bower's (1972) context marker theory. Anderson and Bower suggest that at the time of occurrence of any item there is "simultaneously active a unique element or node in memory [p. 103]" which they call the "list marker" or "list tag." Subsequent differentiation between items in memory is based on retrieving and comparing the list (context) markers associated with the items in question. With reference to the present experiments it would be assumed that when *S* initially encodes the T item he simultaneously encodes contextual information concerning that item. In Experiments I and II *S* first viewed the T items as they were presented to him in the inspection stack. He was seated at a table with *E*, and they viewed the items together. The D items occurred initially in a different context, as a member of a forced-choice pair presented on the WGTA tray. The *E* was not visible when the D items were seen. Thus, the initial appearance of the T items (in a stack, in the presence of *E*) was distinct from the initial appearance of the D items (in a forced-choice pair, in the absence of *E*). These separate contexts could become associated with the two distinct types of items. If, when *S* retrieves an item, he also retrieves this contextual information associated with it at the time of storage, differentiation between T and D items would be possible.

In order to investigate the importance of contextual markers, an attempt was made in Experiment III to minimize contextual differences between the initial presentation of T and D items by presenting the inspection set items in pairs on the WGTA tray. With the exception of the presentation condition the main features of Experiment I were repeated.

Method

Subjects, Apparatus, and Stimulus Materials

These were the same as in Experiment I, with the exception that entirely new pictures were used.

Procedure

The procedure was identical to Experiment I with the following exceptions: (*a*) No additional pretraining was given; (*b*) only 36 forced-choice trials were given each day, corresponding to the three major blocks of trials, 1–12, 13–24, and 27–38, of Experiment I; (*c*) the only major change was the use of the different presentation condition.

The *E* was obscured by the one-way vision screen in all stages of the experiment. The items in the inspection set were initially presented to *S*, not in a stack, but in pairs on the WGTA tray. The *E* placed the pair of items on the tray and presented them to *S* with the instructions, "These are the candy cards." The *S* was told to look at them both and then push the tray back to *E*. The 36 inspection set items were thus presented in 18 pairs on the WGTA tray.

Results and Discussion

The results of Experiment III are shown in Table 1 together with the comparable results of Experiment I. The replication of Experiment I is quite striking. Minimizing obvious contextual differences between T and D items did not lower the high levels of accuracy. These data suggest one of two alternatives: Either context markers are not used to differentiate T and D items, or sufficient contextual information remained to maintain distinct contextual differences between the two types of items. This position illustrates one of the weaknesses of a context marker theory in that the authors (Anderson & Bower, 1972) are quite vague concerning exactly what type of information can serve as context markers. Indeed, one of the main criticisms that can be leveled against the concept of contextual associations is that it merely describes the fact but does not explain how *S* differentiates items in memory (Tulving & Madigan, 1970). Therefore, at this point it seemed inefficient to proceed by systematically isolating other possible contextual cues in an attempt to uncover the specific cues involved. A more economical approach would be to state that the data so far indicated that *S* can tell which items occurred in the inspection set and which ones did not. This global cue, in the inspection set, could comprise many contextual features. Obviously temporal features are also involved. The first occurrence of a T item always preceded that of a D item, and this temporal information could be used to separate the lists. The importance of temporal cues for list discrimination has been documented (Hintzman & Waters, 1969, 1970; Winograd, 1968). Before attempting to separate temporal and contextual elements, it seemed reasonable to test first the validity of the global hypothesis. If, indeed, it is the case that *S* can differentiate T and D items by using the global cue (in the inspection set or not), then subsequent studies could concentrate on the task of isolating which component cues were responsible for performance. In Experiment IV, both T and D items were presented in the inspection set.

EXPERIMENT IV

Method

Subjects

The *S*s were 10 institutionalized retardates selected from the Mansfield State Training School, without regard to sex or diagnostic category. Five of the *S*s had previously taken part in Experiments I, II, and III. The remaining five *S*s were selected from eight *S*s who had previously passed the pretraining criteria described in Experiment I. These five *S*s were selected so that their descriptive statistics would be as close to the five experienced *S*s as possible (mean CA = 16 yr. 2 mo.; mean MA = 9 yr. 11 mo ; mean IQ = 72).

Apparatus and Stimulus Materials

These were the same as in Experiments I, II, and III, with the exception that entirely new pictures were used.

Procedure

All *S*s were tested for a total of 3 days. On each day *S* received two experimental sequences. In each experimental sequence *S* received an inspection set of 36 items, presented to him in two separate stacks of 18 cards. In Condition 1, *S* was given the first stack and told that these were the candy cards. Immediately after viewing the first stack, he was given the second with the instructions that these were not the candy cards. In order to ensure that *S* looked at the negative items, he was required to name the items in both stacks. In Condition 2, the procedure was repeated; however, the initial stack contained the negative items and the second stack consisted of the positive items. Following the viewing of both stacks, the test trials were given. The *S* was always required to choose the candy cards on the test trials, and the position of the correct item was determined by a Gellermann (1933) sequence. The test trials consisted of two trial blocks of 18 forced-choice pairs. In the first trial block, each pair consisted of one T item taken from the positive stack (the candy cards) and one D item selected from the negative stack. In the second trial block, the 18 pairs were represented, but the T and D items were randomly re-paired, following the procedure of Martin (1970). The same stimuli, the T items, were reinforced. Therefore, both frequency and recency of last appearance were controlled on the last trial block.

Each *S* received two experimental sequences each day, one under Condition 1 and one under Condition 2, with the order counterbalanced across *S*s. With 3 days of testing 10 *S*s, the design resulted in 540 observations under each condition.

Results and Discussion

The mean proportions correct on Trial Block 1 were .69 and .68, respectively, for Conditions 1 and 2. On Trial Block 2 the mean proportion correct was .75 for both conditions. A 2 × 2 × 3 × 10 (Conditions × Trial Blocks × Repetitions × *S*s) factorial analysis of variance was conducted on the error scores. The only effect to reach an acceptable level of significance was that due to trial block, $F\ (1,\ 9) = 15.63$, $p < .005$. An inspection of the raw data revealed no difference between naive and experienced *S*s.

Of principal interest here is that performance is about 70% accurate, compared with the 90% levels of accuracy recorded in the preceding experiments. Presenting both types of items in the inspection set drastically reduced the level of performance. The question now arises: What cues associated with the inspection set are used to maintain item differentiation? The possible contribution of temporal cues must now be considered, for in the present series of experiments, the contextual cue "in the inspection set" is confounded with temporal cues.

If it is reasonable to assume that an item's contextual history (Anderson & Bower, 1972) is available to *S*s, it would also be reasonable to assume that its temporal history is also available. Again the distinction must be made between a strength theory and a multiple-trace theory of memory. A strength theory of temporal judgments assumes that a repeating event leaves only one undifferentiated trace in memory. Each additional repetition of an event serves to increase the strength of this trace. Recency judgments are said to be based on the strength of an item's trace in memory. However, if this were the case, it would be impossible to distinguish when the repetitions contributing to that strength occurred. That is, the item's temporal history would be unknown. However, adults cannot only discriminate with some accuracy between the recencies of two independent items (Yntema & Trask, 1963), but they can also distinguish between the recencies of separate occurrences of the same event (Hintzman & Block, 1971). If the temporal history of an item's appearances is available, at least to the extent that *S* can discriminate the recency of more than one repetition of an item (its first appearance from its subsequent appearances), then it would be possible to distinguish T and D items by means of temporal coding. The time-tagging mechanism would run as follows: Items are tagged on their initial appearance. This time tag will be referred to as an item's "birthdate." Subsequent repetitions establish separate traces each tagged with its own recency information. The traces coexist in memory. The birthdates are not obliterated by subsequent repetitions of an item. In the standard condition (Experiments I–III), the birthdates of the T items are always more remote than the D items, and discrimination between them based on temporal cues is possible. However, in Experiment IV, presenting both items in the inspection set reduces the temporal separation between the birthdates, and performance declines.

The concept of "birthdate" may comprise temporal and contextual features in that the item is both time tagged and context tagged when it occurs. An initial attempt to separate context and time features associated with the birthdates was undertaken in Experiment V.

EXPERIMENT V

In this study an attempt was made to examine more closely the levels of performance associated with varying degrees of temporal and contextual distinctiveness associated with the birthdates of the T and D items. Condition 1 represented the standard condition (Experiments I–III), where a clear distinction between the item's birthdates could be made. The T items occurred alone in the inspection set and were followed by forced-choice trials; hence, the D items did not occur until the test trials. In Conditions 2 and 3, both T and D items occurred in the inspection set. However, in Condition 2, the items enjoyed some degree of temporal separation as they were presented in two separate sets, following the procedure of Experiment IV. In Condition 3, the T and D items were presented simultaneously. Thus, Condition 1 represented the greatest degree of temporal separation and an additional context cue (in the inspection set, not in the inspection set). Condition 2 presented a small degree of temporal separation, but both items occurred in the same context. In Condition 3, both items occurred in the same context and the temporal cue was completely removed.

An additional control condition, Condition 4, was also included in this experiment. This was a concurrent discrimination learning situation where *S*'s ability to learn 18 two-choice discrimination problems concurrently was assessed. This condition was included to determine if levels of performance in the experimental groups departed significantly from that which would be expected if *S*s merely learned which item was correct in the course of the experimental sequence.

Method

Subjects

The 10 *S*s consisted of 6 experienced *S*s and an additional 4 *S*s, equated for MA and selected from a pool of *S*s who had met the pretraining requirements described in Experiment I, but had not previously participated in the series. The mean MA of the resulting 10 *S*s was 9 yr. 5 mo., mean CA was 15 yr. 3 mo., and mean IQ was 68.

Apparatus and Stimulus Materials

These were again of the same general description as in previous experiments with all new pictures introduced for the study. The items were preordered into separate sets, one set for each day.

Procedure

Pretraining. The *S*s were tested for a total of 10 days, 2 days of pretraining and 8 experimental sessions. The first day of pretraining consisted of two replications of each of the four conditions of the experiment proper, using inspection sets of 10 items (five pairs). On the second day, this was repeated, using inspection sets of 20 items. The order of conditions was randomized across *S*s. The procedure for each condition was exactly as will be described for the corresponding experimental condition. No success criteria were set.

Experimental sessions. The eight experimental sessions were divided into two replications of each of the four conditions of the experiment. The experimental sequence was divided into three trials. For all conditions on Trial 1, *S* received information concerning which cards were candy cards, and this was indicated by an M&M candy placed directly on top of the correct card. Also, common to all conditions on Trial 1 was that *S*s were required to name each item as it occurred to ensure that they "looked at" the negative items. The procedure was identical for conditions on Trials 2 and 3.

Condition 1: The *S*s originally saw an inspection set of 36 candy cards viewed in pairs on the WGTA, as in Experiment III. Following this viewing period, a forced-choice set of 18 pairs was presented, each pair consisting of one T and one novel D item. Thus, the T item occurred initially in an inspection set and the D items occurred as part of a forced-choice test series exactly as in Experiments I–III. On Trial 2, the remaining 18 cards from the inspection set were paired with the 18 repeating D items, yielding T_1–D_1 pairs. On Trial 3, these 18 pairs were randomly re-paired and presented again, yielding T_2–D_2 pairs.

Condition 2: Trial 1 consisted of a presentation condition similar to Experiment IV. The inspection set consisted of two sets of 18 cards. The sets were presented in pairs on the WGTA tray. The *S*s were told that one set of pairs was the candy cards and the other set was not the candy cards, or vice versa. The order of positive or negative sets was randomized across *S*s. On Trial 2, the 18 pairs for the forced-choice trials were selected so that each pair consisted of one item from the positive set and one from the negative set. The pairs were re-paired randomly for Trial 3.

Condition 3: Trial 1 consisted of an inspection set of 18 pairs presented on the WGTA. One item in each pair was the candy card. On Trials 2 and 3, the 18 pairs were re-paired randomly and presented as forced-choice trials, exactly as in the preceding condition.

Condition 4: In the control condition, 18 concurrent learning problems were presented. On Trial 1, *S* was presented with 18 pairs and told to guess which member of each pair was the candy card. If *S* made an incorrect choice, he was permitted to choose again. If he made a correct choice, an M&M candy was placed directly onto that card. On Trials 2 and 3, the 18 pairs were randomly re-paired, exactly as in the preceding condition.

There were 8 days of testing with one condition occurring each day and two replications of each condition across days. The order of conditions and the assignment of picture sets to conditions were randomized across *S*s. There were two replications of each condition for 10 *S*s. With 18 problems in each of Trials 2 and 3, the design resulted in 360 observations for each trial of each condition.

Results and Discussion

Again, inspection of the raw means revealed no apparent differences between experimentally experienced or naive *S*s, although the new *S*s did appear to have more difficulty during pretraining.

Trial 1

The procedures for Trial 1 were such that Conditions 2 and 3 did not produce data, as *S*s merely viewed the items. In Condition 1, however, *S*s received a standard forced-choice recognition test as part of Trial 1. Performance at distinguishing the old (T item) from the novel distractor (D item) was 93% correct. In Condition 4, *S*s were required to choose between the items in a pair on a chance basis. The proportion correct was .49.

Trials 2 and 3

The mean proportion correct as a function of condition and trial is shown in Table 3. Performance is considerably more accurate in Condition 1 than in the remaining conditions.

As no effect of repetitions was apparent in the raw means, the data were collapsed across this variable. A 4 × 2 × 10 (Conditions × Trials × *S*s) factorial analysis of variance was conducted on the error scores. The main effect of conditions was highly significant, *F* (3, 27) = 214.54, $p < .001$. This effect was followed up according to the procedures outlined by Scheffé (1959). All pair-wise differences were reliable, S^2 (3, 27) = 9.12, $p < .05$.

The main effect of trials was also significant, *F* (1, 9) = 87.89, $p < .001$, as well as the Conditions × Trials interaction, *F* (3, 27) = 4.03, $p < .025$. This interaction was due to the fact that there was a reliable increase in performance over trials in Conditions 2–4, S^2 (3, 27) $\leq$ 25.81, $p < .001$, whereas there was no effect due to trials in Condition 1, S^2 (3, 27) = 1.61, $p > .10$. The lack of a significant increase in accuracy for Condition 1 possibly represents a ceiling effect.

TABLE 3

Proportion of Correct Responses on Trials 2 and 3 as a Function of Condition, Experiment V

Trial	Condition			
	1	2	3	4
2	.89	.66	.62	.52
3	.91	.76	.71	.62

Note. All proportions are based on 360 observations.

Considering the three experimental conditions, where the degree of temporal separation between the T and the D items is varied, it is noticeable that the magnitude of the difference between Condition 1 and the other two conditions (2 and 3) is quite large. The average accuracy level of 90% replicates preceding experiments for this condition where the T items alone are presented in the inspection set. When both items occur in the inspection set, performance drops to approximately 70% correct, again replicating preceding findings. Thus, it appears to be the case that providing distinct contextual and/or temporal cues, which can be associated with an item's birthdate, leads to excellent performance. Of interest is the fact that the difference between 71% correct for Condition 2 and 66% correct for Condition 3 does represent a reliable difference. Thus, even a slight temporal separation provided by presenting all of one type of item first, within the same context, does help later discrimination between them, although this increase is small.

As a final point it is interesting to note that performance on the concurrent learning problems is considerably less efficient than in the other conditions, with a mean performance level of 57% correct. These data suggest that the procedure of indicating which item is correct on an inspection trial, before *S* makes his choice, results in a significantly higher level of accuracy. Consider Conditions 3 and 4. Under both conditions *S* is presented with a pair comprised of a positive and negative item. In Condition 4 *S* chooses the candy card by chance but is allowed to correct incorrect choices. Thus, he does have the necessary information concerning which is the candy card on each trial. In Condition 3, the main difference is that *S* never chooses an incorrect card. He is told immediately which card is correct. Thus, the tag of "candy card" or "not candy card" can occur simultaneously with the initial encoding in Condition 3, but in Condition 4 must await feedback. Further investigation of the importance of remembering not only which item is chosen but whether it was correct or incorrect may lead to interesting data concerning the effects of memory in discrimination learning (Buchwald, 1969).

GENERAL DISCUSSION

These data have important implications for both theories of memory in general and for theories of retardate memory in particular. The implications for general theories will be considered first.

Neither a traditional strength theory nor a frequency attribute theory of recognition memory can readily explain the present data. Equating the experimental frequency and recency of all items should also equate their memory strength. Nonetheless, T and D items are still differentiable. Similarly, that accuracy is not lowered when the frequencies of occurrence were equated presents difficulties for frequency attribute theory (Underwood & Freund, 1970).

The relation of the present data to Anderson and Bower's (1972) context marker theory is more complex and brings into focus a weakness in both the theory and in the procedure of these studies. Manipulating obvious contextual elements associated with the presentation of T and D items (Experiment III) did not influence performance. This may be because context markers were not involved or the context elements manipulated were not the prime determinants of performance. The weakness of the theory is that it does not specify which context elements are likely to be used as markers. Only trial and error with a particular task would determine this. The resultant argument is, of course, circular. If varying context elements influences performance, context markers are involved, and the theory supported. If such manipulations do not change performance, the appropriate context cues have not been tapped.

The weakness of the procedures used in these studies is that it is not possible to entirely separate contextual and temporal cues, for the birthdate of an item may be a combination of temporal and contextual features. Isolating the T items in the inspection set provided both the greatest degree of temporal separation between the birthdate of the T and D items and an additional set of distinct contextual associations. The T items occurred in a separate context, the inspection set. Thus, the resultant birthdate information is a combination of temporal and contextual features. However, the present data do suggest that items in memory are stored together with list markers which convey information concerning their past histories. This historical information consists of temporal and/or contextual associations. If an item is repeated, each repetition will be associated with a separate set of list markers. The separate list markers can be discriminated in memory at least to the extent that information associated with an item on its first occurrence (its birthdate) is not obliterated by subsequent repetitions.

Of interest to developmental theories is the excellent level of performance reported for retarded *S*s, a population which has been traditionally described as deficient in memory processes. However, recognition memory for pictures is very efficient for these children. The probability of identifying an old item under the standard recognition memory procedure remained at approximately .96 over a wide variety of experimental manipulations.

One possible explanation for the excellent performance is that the tasks reported here are concerned with long-term memory (LTM), and while retardates have been shown to possess a short-term memory (STM) deficit (Scott & Scott, 1968), clear evidence of an LTM deficit is lacking (Belmont, 1966). An alternative, or possibly an additional, explanation is that retardates, like young children, are deficient in memory tasks requiring deliberate mnemonic strategies to mediate performance (Flavell, 1970). It is not clear that the modified recognition memory tasks described here involve any such deliberate strategy on the part of *S*. Temporal information can be stored even when *S*s are not warned that they will be asked for such information

(Hintzman & Block, 1971) and without any apparent "cost" to the memory system (Zimmerman & Underwood, 1968). Thus, it appears that the storing of temporal information regarding items is not necessarily under *S*'s control and does not, therefore, involve a deliberate mnemonic strategy on the part of *S*. Whether the distinction is between STM and LTM tasks, or between tasks requiring or not requiring mnemonic strategies, remains a subject of further research. However, these studies confirm that retarded *S*s are extremely efficient not only at recognizing that they have seen an item before but also where and when that item occurred.

REFERENCES

ANDERSON, J. R., & BOWER, G. H. Recognition and retrieval process in free recall. *Psychological Review*, 1972, **79**, 97–123.

BELMONT, J. M. Long-term memory in mental retardation. In N. R. Ellis (Ed.), *International review of research in mental retardation.* Vol. 1. New York: Academic Press, 1966.

BROWN, A. L. Context and recency cues in the recognition memory of retardates. *American Journal of Mental Deficiency*, 1972, **77**, 54–58.

BROWN, A. L., & CAMPIONE, J. C. Recognition memory for perceptually similar pictures in preschool children. *Journal of Experimental Psychology*, 1972, **95**, 55–62.

BROWN, A. L., & SCOTT, M. S. Recognition memory for pictures in preschool children. *Journal of Experimental Child Psychology*, 1971, **11**, 401–412.

BUCHWALD, A. M. Effects of "right" and "wrong" on subsequent behavior: A new interpretation. *Psychological Review*, 1969, **76**, 132–143.

CAMPIONE, J. C. Transitivity and choice behavior. *Journal of Experimental Child Psychology*, 1969, **7**, 387–399.

FLAVELL, J. H. Developmental studies of mediated memory. In H. W. Reese & L. P. Lipsitt (Eds.), *Advances in child development and behavior.* Vol. 5. New York: Academic Press, 1970.

GELLERMANN, L. W. Chance order of alternating stimuli in visual discrimination experiments. *Journal of Genetic Psychology*, 1933, **42**, 207–208.

HINTZMAN, D. L., & BLOCK, R. A. Repetition and memory: Evidence for a multiple-trace hypothesis. *Journal of Experimental Psychology*, 1971, **88**, 297–306.

HINTZMAN, D. L., & WATERS, R. M. Interlist and retention intervals in list discrimination. *Psychonomic Science*, 1969, **17**, 357–358.

HINTZMAN, D. L., & WATERS, R. M. Recency and frequency as factors in list discrimination. *Journal of Verbal Learning and Verbal Behavior*, 1970, **9**, 218–221.

KINTSCH, W. Models for free recall and recognition. In D. A. Norman (Ed.), *Models of human memory.* New York: Academic Press, 1970.

MARTIN, A. S. The effect of the novelty-familiarity dimension on discrimination learning by mental retardates. Unpublished doctoral dissertation, University of Connecticut, 1970.

SCHEFFÉ, H. *The analysis of variance.* New York: Wiley, 1959.

SCOTT, K. G., & SCOTT, M. S. Research and theory in short-term memory. In N. R. Ellis (Ed.), *International review of research in mental retardation.* Vol. 3. New York: Academic Press, 1968.

TULVING, E., & MADIGAN, S. A. Memory and verbal learning. *Annual Review of Psychology*, 1970, **21**, 437–484.

TULVING, E., & THOMSON, D. M. Retrieval process in recognition memory: Effects of associative context. *Journal of Experimental Psychology*, 1971, **87**, 116–124.

UNDERWOOD, B. J., & FREUND, J. S. Testing effects in the recognition of words. *Journal of Verbal Learning and Verbal Behavior*, 1970, **9**, 117–125.

WINOGRAD, E. List differentiation as a function of frequency and retention interval. *Journal of Experimental Psychology Monograph Supplement*, 1968, **76**(2, Pt. 2).

YNTEMA, D. B., & TRASK, F. P. Recall as a search process. *Journal of Verbal Learning and Verbal Behavior*, 1963, **2**, 65–74.

ZIMMERMAN, J., & UNDERWOOD, B. J. Ordinal position knowledge within and across lists as a function of instructions in free-recall learning. *Journal of General Psychology*, 1968, **79**, 301–307.

(Received August 3, 1972)

Journal of Experimental Psychology
1973, Vol. 98, No. 1, 14–24

REPRESENTATION OF SPEECH SOUNDS IN PRECATEGORICAL ACOUSTIC STORAGE[1]

ROBERT G. CROWDER[2]

Yale University

Three experiments are described in which *Ss* gave immediate serial recall for lists of seven acoustically presented synthetic syllables. In the first study the to-be-remembered information was contained in terminal voiced stop consonants of vowel–consonant syllables, while in the second two studies the to-be-remembered information was contained in initial voiced fricatives of consonant–vowel syllables. There were statistically significant recency effects and stimulus suffix effects for the fricatives but not for the stops. These results pertain to earlier findings contrasting memory performance on stop consonants to performance on vowels. The suggestion is that this vowel–consonant difference results neither from the position which the critical information occupies in a syllable nor from mechanisms correlated with lateralized or categorical perception of speech; rather it appears to be crucial whether the discrimination is based on steady-state or transient information. Although this conclusion is consistent with the data, it must remain somewhat tentative since (*a*) the absolute magnitude of recency and suffix effects was small with fricatives, and (*b*) it cannot be assumed that fricatives are inevitably perceived in the same (e.g., categorical) manner.

In a recent report (Crowder, 1971a) I showed that immediate memory for vocabularies of three consonant–vowel (CV) syllables is quite different depending on whether the to-be-remembered information is contained in the initial voiced stop consonants, as in /ba, da, ga/, or in the terminal vowels, as in /ga, gæ, gʌ/. When the *terminal vowels* of CVs are being remembered, as in the latter vocabulary above, there is a marked recency effect with auditory presentation (advantage in recall of the last serial position over the second-to-last serial position) and a significant effect of a stimulus suffix (Crowder & Morton, 1969) upon the last part of the series. Both the recency and suffix effects are characteristic of natural stimuli such as digits and letters as well as of CV syllables with varied vowels. When the *initial stop consonants* of CVs were to be remembered, on the other hand, there was neither a significant recency effect nor a suffix effect. Furthermore (Experiment V), direct comparisons of visual and auditory presentation revealed the usual advantage of auditory presentation for a vocabulary with varied terminal vowels, but no modality effect at all for a vocabulary with varied initial stop consonants. These findings of different immediate memory for vowels and consonants are consistent with reports by Cole (1973), Smallwood and Tromater (1971), and Fujisaki and Kawashima (1970).

One theory which can handle the modality effect (auditory better than visual presentation), the recency effect (last position better than next-to-last with auditory presentation), and the suffix effect (detrimental effect of a redundant word on final serial positions of an auditory series) has been offered by Crowder and Morton (1969). This theory, most recently elaborated in Morton, Crowder, and Prussin (1971), states that following auditory input of a memory list there is information in sensory form (precategorical acoustic storage [PAS]), which remains available for at least 2 sec., concerning the last few posi-

---

[1] This research was supported in part by National Science Foundation Grant GB15157. Thanks are due Haskins Laboratories, New Haven, Connecticut, for allowing the author to use their facilities in preparing stimuli; also the author wishes to thank Henry L. Roediger, III, for testing *Ss*.

[2] Requests for reprints should be sent to Robert G. Crowder, Department of Psychology, Yale University, 333 Cedar Street, New Haven, Connecticut 06510.

tions. The availability of PAS information concerning the sounds of the last few items accounts for the advantage of auditory over visual input and for the related observation of increased performance on the last items in the auditory case (recency). It is assumed that the fixed-capacity PAS system is subject to displacement effects, such as the disruptive effect of a stimulus suffix. Since the system is precategorical (with regard to acquired linguistic classifications) it makes no difference whether or not the suffix word is from a different class than the memory materials (see Morton et al., 1971).

Apparently the PAS hypotheses (or any hypotheses aimed at the effects of auditory presentation), while adequate for natural stimuli and for tasks requiring memory for terminal vowels, are superfluous when it comes to initial stop consonants since the three symptoms of acoustic memory (the modality, recency, and suffix effects) are absent in tasks requiring memory for the stops. A tentative conclusion could thus be drawn that not all speech sounds are stored in PAS, as would have been the case with a "tape recorder" memory, but rather only certain parameters of the acoustic stream are retained in this echoic form.[3] Since natural-language stimuli are quite redundant, a memory system such as PAS, holding only certain parameters of the signal, would still be helpful in disambiguating perceptual or memorial confusions.

In the present research I have tried to begin identifying which aspects of the initial stop consonants and the terminal vowels used earlier were responsible for the differences obtained. Note that the obtained contrast between these two types of stimuli is interesting regardless of whether the PAS theory is correct. Whatever mechanism or mechanisms produce the symptoms of acoustic memory *must* be differentially active for different kinds of speech sounds. The following three possibilities concerning why stop consonants gave surprising results in the earlier study were put to test:

1. The difference could have been produced by the confounding of letter class with position in the CV syllable, since the stop consonants were always at the beginning of each syllable and the vowels at the end. If PAS or some other mechanism were sensitive to the last piece of information in each segment the result would have been an apparent vowel–consonant difference. Experiment I of the present series rules out this possibility by showing that terminal stop consonants resemble the initial stops examined previously.

2. The difference could have been produced by the confounding of letter class with the extent to which the discriminating cue varied over time. The cues for stop consonants are (complex) rapid frequency sweeps, largely in the second formant, whereas the cues for vowels are static relations between the formant frequencies. If PAS or some other mechanism were differentially sensitive to *steady-state information*, the obtained result would have occurred. The second and third experiments of the present series suggest that this is indeed the case since results obtained with a vocabulary of steady-state consonants, fricatives, were quite similar to those obtained with vowels.

3. It is possible that the original vowel–consonant difference could have been produced by an inherent relationship between letter class and special mechanisms of speech perception. One mark of such a special mechanism is found in research on laterality differences in dichotic speech recognition (Shankweiler & Studdert-Kennedy, 1970), showing a right-ear advantage for stop consonants, a left-ear advantage for musical sounds, and symmetrical accuracy for vowels. Since auditory projections are stronger contralaterally than ipsilaterally the right-ear advantage is consistent with localization of speech and language functions in the dominant (usually left) hemisphere of the brain. It could be that PAS is thus a property of the nondominant hemisphere, where vowels are uniquely or largely handled, and so vocabu-

---

[3] Of course there must in principle be sensory memory for all speech signals; the question at issue is whether relatively long-lasting representation is differentially available to different classes of speech sounds.

laries which show the right-ear advantage should not give evidence of PAS. A second mark of special speech-processing mechanisms is the phenomenon called categorical perception. Liberman, Mattingly, and Turvey (1972) have applied notions of categorical perception to my report of the absence of PAS in stop consonants as follows:

> at least in general terms, then, the difference in recency between the stops and vowels is exactly what we would expect. As we have seen, . . . the special process that decodes the stops strips away all auditory information and presents to immediate perception a categorical linguistic event the listener can be aware of only as (b, d, g, p, t, k). Thus, there is for these segments no auditory, precategorical form that is available to consciousness for a time long enough to produce a recency effect. The relatively unencoded vowels, on the other hand, are capable of being perceived in a different way. Perception is more nearly continuous than categorical . . . because the auditory characteristics of the signal can be preserved for a while [p. 329].

Liberman et al. are thus proposing that since even the perception of stop consonants gets carried out with only minimal acoustic guidance it is no surprise that acoustic information does not linger in memory afterward.

Although these two related hypotheses (laterality and categorical perception) are rather clearly more interesting than Possibility 1 or 2 above, the evidence seems to argue against them since fricatives (used as stimuli in Experiments II and III) both show the features of special speech processing (Darwin, 1971) and also give memory results like vowels (which do not receive such special processing ordinarily).

Experiment I

In the first experiment the seven-syllable series were randomly composed of elements from a vocabulary of three synthetic-speech VC syllables, /Ib, Id, Ig/. If the earlier consonant–vowel differences (Crowder, 1971a) came from the position of vowels at the end of CV syllables in that study, then in the present experiment, where the stop consonants occupy these favored terminal positions, there ought to be evidence for significant recency and significant suffix effects with auditory presentation.

Method. Insofar as possible the stimuli were comparable to those used in the earlier study except for the change in position of the stop consonant and the change in vowel from /a/ to /I/ (which made synthesis of convincing terminal stops easier). The syllables were first generated according to Mattingly's (1968) Speech Synthesis by Rule routines[4] and then altered so as to be 300 msec. in length. The intelligibility of these syllables suffered since it was not allowable to provide releases of the terminal stops (e.g., "ih-buh, ih-duh, ih-guh") without making the VC syllables into VCVs (see Mattingly, Liberman, Syrdal, & Halwes, 1971). Data presented below describe the obtained level of intelligibility and indicate that it was not a factor in the results of the memory trials.

The *S*s were 26 college-age paid volunteers with summer residence in the New Haven area; they had responded to signs posted around Yale University campus buildings. After an introductory description of the experiment they were played numerous examples of the three VC syllables and then given a 28-item intelligibility test wherein each syllable was played alone (in haphazard order) and 1 sec. allowed for written identification. Here, and in the subsequent memory experiment, *S*s were told simply to use the letters *b*, *d*, and *g* to stand for the three VC syllables.

All *S*s were run in groups varying in size from one to five. There was a fixed list of 50 seven-syllable series composed from a table of random numbers; these were exactly the same stimuli which I had used earlier (Crowder 1971a, Experiments I–IV) with the syllable /Ib/ substituted for /ba/, /Id/ for /da/, etc. The session was divided into five 10-trial blocks. In the first, third, and fifth trial blocks, the seventh syllable in the memory series was followed by a 1,100-Hz. tone. The presentation rate of the series was two syllables/sec, the tone occurring in time with this rate. On the second and fourth blocks of trials, the spoken word "zero" was substituted for the tone.

Results. The second panel of Figure 1 shows error rates as a function of serial position separately for the control (tone) and suffix (zero) conditions. Although there was an apparent superiority of the final position over the next-to-last position in the control condition, this difference was slight in magnitude and did not approach statistical significance as measured by a Wilcoxon test, $T = 126$, $z = .68$. Thus, as with initial stops, a recency effect was not obtained with terminal stops under condi-

---

[4] The author appreciates the assistance of James E. Cutting in developing the actual syllables used.

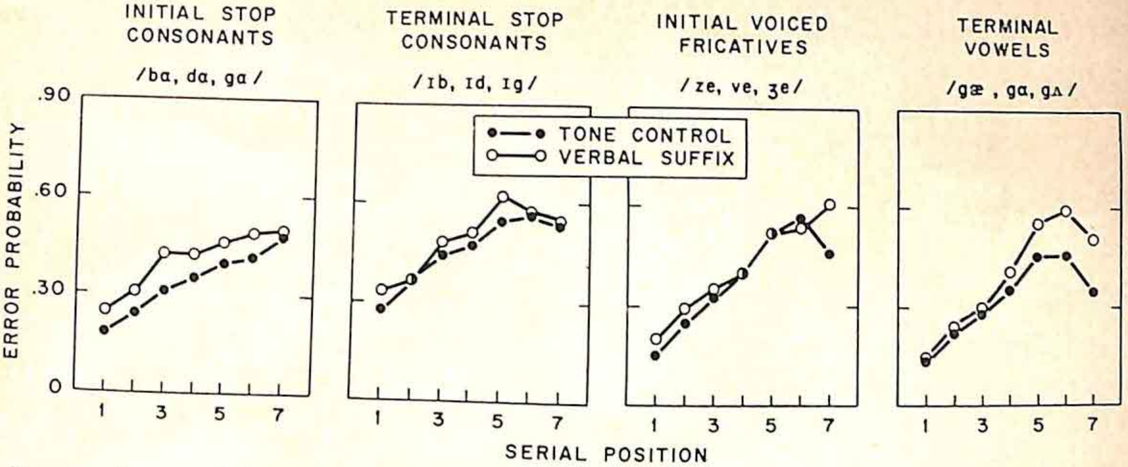


FIG. 1. Error probability as a function of serial position. (In each panel a control condition [tone following last item] is compared with a verbal suffix condition [redundant word following last item]. The data plotted in the first and fourth panels are taken from Crowder, 1971a, for comparison with the present data, which are shown in the second and third panels.)

tions where a tone followed the last item. The measure of the suffix effect used was the difference between suffix and control conditions on the last serial position. The results of a Wilcoxon test again confirmed the similarity of results for initial and terminal stop consonants; i.e., no statistically significant advantage of the control condition over the verbal suffix condition was obtained, $T = 165$, $z = .38$.

Table 1 is presented to show that the outcome regarding terminal stop consonants was *not* a function of the overall task difficulty. The results of the 28-item intelligibility pretest indicated it was quite hard to understand the syllables even in isolation; the median error rate was between 9% and 10%. (However, practice seemed to help and if intelligibility had been sampled later during the experiment scores would probably have been higher.) The memory scores of many *S*s approached chance performance at the more difficult serial positions. If an individual were deaf, his score on a block of 10 trials through guessing alone would be, on the average, .67 × 10 (Trials) × 7 (Positions) = 46.9, with a standard deviation of 3.9. It was decided arbitrarily to isolate two subgroups of the total *N* of 26: those who came within the 5% confidence limit (a score of 40.5 errors) of chance performance on *any* block of 10 trials as opposed to those who consistently scored below this level of errors. As it turned out there were 12 *S*s in the poor-scoring group and 14 in the better-scoring group, and, as Table 1 shows, neither subgroup violated the conclusions reached on the basis of all the data. Though there appeared to be an overall

TABLE 1

ERROR PROBABILITY AS JOINT FUNCTIONS OF RECALL CONDITION AND SERIAL POSITION FOR DIFFERENT SUBGROUPS OF *S*s

Serial position	Subgroup of *S*s		
	Overall high performance ($n = 14$)	Overall low performance ($n = 12$)	Largest "recency" in control ($n = 5$)
	Tone control		
1	.24	.40	.31
2	.28	.49	.41
3	.40	.48	.51
4	.43	.51	.52
5	.52	.60	.62
6	.55	.57	.67
7	.52	.55	.43
	Zero suffix		
1	.28	.41	.25
2	.29	.50	.34
3	.42	.55	.45
4	.48	.56	.56
5	.59	.67	.66
6	.53	.61	.58
7	.50	.60	.49

(nonselective) effect of the suffix on *S*s with relatively poor scores this was not statistically significant (Wilcoxon $T = 30.0$).

Finally, in Table 1 a separate subgroup of five *S*s is shown; these cases were selected on the basis of having the largest "recency" effect in the control condition. It could be argued that a suffix effect was not found in the present study simply because there was no recency in the control condition *to be removed* by the suffix. If this were so, taking only those *S*s who happened (by chance?) to have shown a recency-like advantage of the last position over the next-to-last would have provided an opportunity to see the suffix effect at work. However, as Table 1 makes clear, these *S*s revealed no suffix effect such as can be seen for vowels or in the second experiment to be reported below.

Discussion. It is always dangerous to be left believing the null hypothesis because unless an explicit single alternative is being entertained one cannot assign a probability estimate to its truth. For that reason substantially more *S*s were tested here in Experiment I than are usually necessary for studies of similar design. Yet it was not possible to squeeze from the data either a reliable advantage for the last serial position or a reliable suffix effect, even when *S*s were divided into potentially interesting subpopulations.

It appears that stop consonants (at least the voiced stops) differ from other memory stimuli in whatever properties govern occurrence of the recency and suffix effects in vowels or in natural characters. The present study shows that this is true whether the vowel environment follows the consonant or comes before it. Accordingly, the earlier findings on classes of speech sounds must not be an artifact of the relative positions occupied by the to-be-remembered information in the syllable. We can thus turn to more interesting hypotheses for the vowel–consonant difference.

EXPERIMENT II

The remaining possibilities for explaining the vowel–consonant difference are listed as Possibilities 2 and 3 above, viz., explanations based on transient vs. steady-state cues or based on the perception and primary representation of consonants through special perceptual mechanisms in the dominant hemisphere and of vowels in the nondominant hemisphere. It would be very difficult to test between these notions if all consonants had transient acoustic cues and were also known to be lateralized in the dominant hemisphere (i.e., showed the right-ear effect in dichotic tasks). Fortunately, there are classes of letters, particularly the *fricatives*, where these two characteristics of stop consonants are dissociated. The fricatives are distinguished by steady-state information, like vowels, but show a sizeable right-ear advantage, like stop consonants when placed in CV syllables with transitions to the terminal vowel (Darwin, 1971). Unfortunately, data relevant to the *categorical perception* outcome are not available for Darwin's fricatives. In other vocabularies, categorical perception tends to occur for stimuli leading to a right-ear effect but not for those which do not, although there are exceptions. In any case, these considerations argue for couching the present logic in terms of laterality rather than in terms of categoricality. Thus, if vocabularies of fricative stimuli produce a recency effect and suffix effect in immediate memory then the transient vs. steady-state distinction would be supported in the vowel–consonant difference. If fricatives showed memory results like stop consonants then the special-processing (laterality) hypothesis would be supported.

Method. The method in Experiment II was similar to that used in Experiment I with the exceptions noted. The stimulus vocabulary consisted of the three voiced fricatives /ve, ze, ʒe/ in 300-msec. CV syllables. These were the same stimuli (i.e., same parameters and same synthesizer) as used in Experiment I of Darwin (1971) except that the length of steady-state friction was extended from his 45 msec. to 100 msec., the terminal consonant used in his study was removed, and the total length was then set at 300 msec. by adjusting the duration of the vowel sound. It was important to use exactly these stimuli since Darwin demonstrated a marked right-ear advantage for them in a dichotic identification task.[5]

The stimuli and their order were exactly the same as in the first study, except for one-to-one substitution of the items /ve/ for /ɪb/, /ze/ for /ɪd/, and /ʒe/ for /ɪg/. The *S*s were allowed to use any recall symbol they chose for the /ʒe/; how ever, *s* was suggested by *E* and most *S*s used it. Sixteen *S*s participated; all had also been participants in Experiment I.

---

[5] These stimuli included formant transitions.

Results. The main results are given in the third panel of Figure 1. It is an easy outcome to describe—there was a highly significant recency effect from the next-to-last to the last serial position, $T = 23.5$, $z = 2.30$, $p = .01$, and a highly significant suffix effect, $T = 22.0$, $z = 2.38$, $p < .01$, which was specific to the last position. The statistical reliability of these results is quite remarkable considering the moderately small numerical differences and the fact that only 16 *S*s were tested, rather than the 26 used in the first study. Intelligibility of the syllables in isolation was considerably better here than in the first study; only six *S*s made any errors at all and only two of these made more than one error on the 28-item pretest. This greater ease of identification is quite unlikely to have anything to do with the differential outcomes of Experiments I and II, however, since no recency or suffix effect was found, even for *S*s in Experiment I who found the task easy, and since performance on Position 6 in the two studies was almost exactly comparable in the control condition.

Discussion. The first and fourth panels of Figure 1 are from the earlier experiment and are included for purposes of comparison. Note that whereas the spoken word "zero" was the verbal suffix used in the present two studies, single syllables were used in the earlier two; however, the tone control conditions were exactly comparable.[6]

The evidence in Figure 1 permits the conclusion that vowels and consonants in the fourth and first panels gave contrasting results mainly because the vowels are distinguished on the basis of steady-state acoustic cues while the consonants are distinguished on the basis of dynamic formant transitions. One of the next steps will be to show that other sounds which are not consonants, but which depend on formant transitions (such as perhaps /eI/ and /aI/), show results like stop consonants.

EXPERIMENT III

The final study was intended to establish additional reliability and generality to the result obtained above with fricative stimuli; the justification for this effort is the considerable theoretical importance which has been attached to the outcome with fricatives.

In the new experiment, some *S*s were tested with fricative stimuli in which no terminal vowels were provided whereas the other *S*s heard stimuli consisting of fricative–vowel syllables, much like those used in Experiment II. The idea behind using these two versions of fricative items was twofold. First, one is always nervous about research with synthetic speech for fear the particular way sounds are realized on the computer might have more to do with the results than the properties of those sounds per se. Second, Darwin (1971) found important differences between the lateralization of fricative perception depending on whether terminal vowels (and their initial formant transitions) were included with the fricatives or not. In particular, Darwin showed a right-ear advantage, in dichotic identification when fricatives were followed by formant transitions and vowels (i.e., with the stimulus items used in Experiment II of the present series), which was comparable in magnitude to the right-ear advantage shown by the same *S*s for sounds identified as /b, d, g/. However, when essentially the same *fricative* sounds were presented but without terminal vowels (and therefore also without transitions into these vowels), no right-ear advantage was obtained.

The fact that fricative–vowel syllables are lateralized while isolated fricatives are not permits a converging test on the major point raised in Experiment II—whether the original (Crowder, 1971a) difference between memory for vowels and stop consonants is based on consonants being transient and vowels being steady state or whether this difference is based on consonants being selectively processed in the dominant hemisphere and vowels not. If the laterality hypothesis were true then isolated fricatives should yield results more

---

[6] Since different suffixes were employed in these studies not much should be made of the apparently greater difference between experimental and control conditions in the fourth panel (terminal vowels) than in the third panel (initial voiced fricatives).

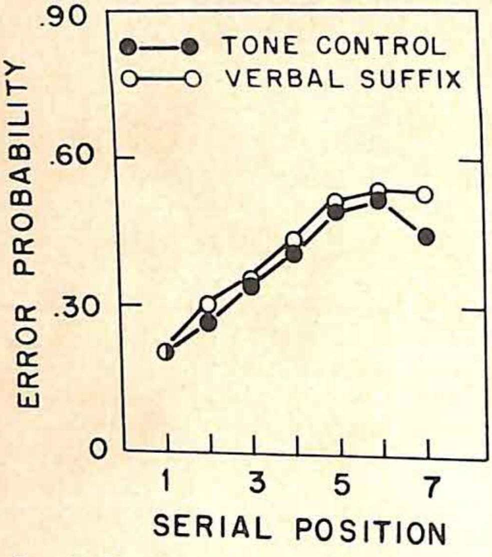


FIG. 2. Error probability as a function of serial position for the combined fricatives in Experiment III.

like vowels (i.e., large recency and suffix effects) than fricatives with terminal-vowel environments. This is because in perception, as noted in the previous paragraph, isolated fricatives are also more like vowels (i.e., no lateralization) than are fricatives with vowels.

Method. Thirty-two *S*s (from the same source as used previously but naive to the particular type of experimental situation under study) received 60 trials on which memory for seven-item serial lists was tested. All *S*s heard the same lists in the same order; however, for 16 *S*s all 60 trials contained isolated fricative stimuli and for the other 16 *S*s all 60 trials contained fricatives with a redundant vowel (always /e/). For half of each group, the first 30 trials used stimulus series followed by a tone (same as in Experiment II) and for the second 30 trials the spoken word "zero" was used; the order of these conditions (control and experimental) was reversed for the other half of the *S*s. Thus, the experimental design was a 2 × 2 factorial, between-*S* comparisons on the first factor (presence or absence of vowels in the to-be-remembered items) and within-*S* comparisons on the second factor (suffix vs. control).

The stimulus syllables, as in Experiment II, were based on those used by Darwin (1971). The fricative–vowel items were 295 msec. long, with 125 msec. of friction, 30 msec. of formant transition, and 140 msec. of the terminal vowel /e/. In addition the amplitudes of the friction segments were raised uniformly so as to increase intelligibility. The isolated fricative items were simply the initial 125 msec. of the fricative–vowel syllables. Other details of Experiment III were comparable to those of Experiment II. In summary, the present design extends that of the previous study by using fricative stimuli both *with and without* formant transitions and vowels, *S*s working with one or the other version of these fricatives, suffix and control observations being taken in two balanced blocks of 30 trials.

Results. The main result is given in Figure 2, which shows performance summed across the two types of fricative items Although neither the recency effect in the control condition nor the effect of the suffix on the last position seems quite as large as in the third panel of Figure 1, the pattern of findings is identical and both outcomes are highly significant statistically for the recency effect, $T = 103.0$, $z = 2.28$ $p = .01$; for the suffix effect, $T = 105.5$ $z = 2.42$, $p < .01$.

Verifying the fricative results from Experiment II was the main purpose of the present study and, as Figure 2 shows, this replication was successful. However, the design also permits a comparison between two versions of fricative items—those with terminal vowels and those without. Recall that this comparison could bear on the major theoretical issue at stake here in that isolated fricatives are distinguished from fricative–vowel syllables in lateralized perception (Darwin, 1971); specifically, the isolated fricatives should give stronger recency and suffix effects than the CV syllables if the lateralization of speech processing is an important dimension in distinguishing the original results for vowels and consonants in immediate memory

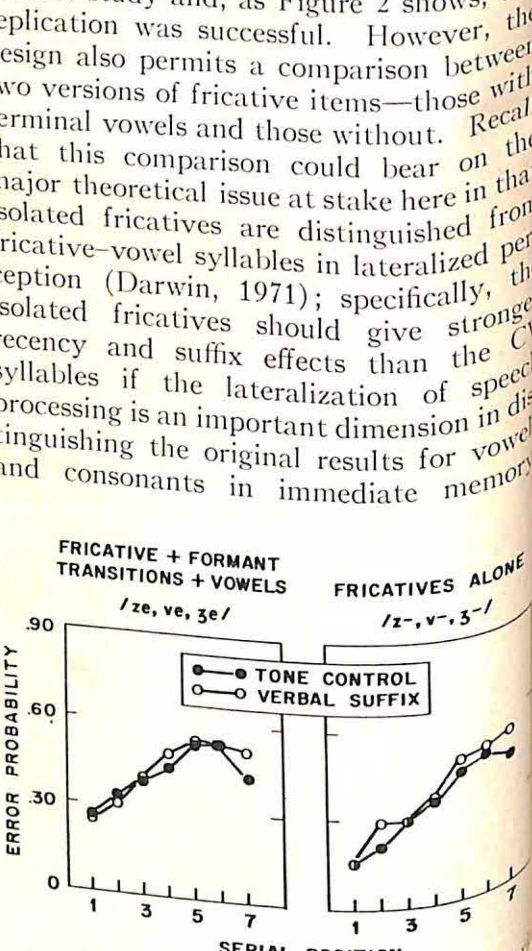


FIG. 3. Error probability as a function of se position for isolated fricatives and fricative–vo syllables in Experiment III.

Shown separately in Figure 3 are the results for the fricatives with redundant terminal vowels and for those without vowels. Each panel is based on 16 independent *S*s, serving 30 trials under each of the two conditions. To put it briefly, there is no suggestion either in the figure, or in the statistical analyses associated with it, that the isolated fricatives produced larger recency or suffix effects than the fricatives which blended into redundant vowels. The apparently larger recency effect between Positions 6 and 7 in the fricative–vowel condition was not a statistically reliable interaction.

There was quite a difference between the two types of stimuli in the initial 10-item intelligibility test: Four of 16 *S*s in the fricative–vowel group made more than one error whereas 11 of 16 in the fricative-alone group made more than one error, $\chi^2 = 143.04$, $z = 11.9$, $p < .01$. This advantage of the CV syllables over the isolated friction is difficult to reconcile with the substantial advantage the latter had over the former on the first few serial positions of the memory experiment (see Figure 3). Perhaps hearing fricatives in close succession makes discrimination easier by converting an absolute to a relative judgment task. If so, this context effect would work especially well when no vowels separate successive fricatives.

Discussion. The failure to find larger recency and suffix effects with isolated fricatives than with fricatives in vowel context is not consistent with the hypothesis that other consonants, viz., voiced stops, fail to show these effects because they are processed in the speech-dominant hemisphere. It could be argued that the degree of lateralization found by Darwin (1971) to differentiate fricatives with and without transitions to vowels was rather slight; however, he found the right-ear advantage for stop consonants to be similarly small in magnitude. The fact that stop consonants and fricatives with terminal vowels are processed at an advantage in the left hemisphere (and that vowels and isolated fricatives are not) seems, in light of this experiment, to have nothing to do with the failure of stop consonants to reveal significant recency and suffix effects in immediate memory.

MEMORY FOR FRICATIVES: COMMENT

There is no doubt that by the conventional rules of evidence in psychology fricatives indeed show performance patterns theoretically attributed to PAS. As argued above, that in turn suggests that such patterns are independent of the peculiar manner in which some speech sounds are processed perceptually (categorically, in the dominant hemisphere). Still, the Liberman et al. (1972) argument—that speech sounds failing to show PAS are the same sounds which are stripped of their acoustic correlates very early in the course of categorical perception—has tremendous allure logically. The perception of stop consonants defies such everyday psychophysical laws as the superiority of relative over absolute judgments. That is, with categorically perceived stop consonants *S*s cannot hear differences in acoustic stimuli unless those differences are commonly denoted by different names (Pisoni, 1971, gives some of the most recent and strongest evidence for this striking result). Assuming a continuity between experiments on the perception of speech and experiments on the immediate retention of those same sounds, it would have been paradoxical if stops *had* shown evidence for PAS. If stops had enjoyed a precategorical code in memory why wouldn't the same code be available in perception experiments where sometimes the only task is to decide whether two successive sounds are identical. That is, if the functional equivalent of an echo is available in memory for the most recently presented item (essentially the claim made by PAS theory) then why shouldn't the same echo be effective in improving same–different scores beyond predictions based on assignment of names to the sounds.

But granted the important similarities between fricatives and stops (similar right-ear advantage, for example) we are left with exactly the paradox described, evidence for unclassified acoustic information in the memory paradigm but not in the perception paradigm (provided fricatives are perceived categorically). For these reasons the conclusions reached on the basis of Experiments II and III deserve closest scrutiny.

There might be dangers related to the particular stimuli used. Several people involved in the fricative studies noted that of the three voiced fricatives used, /z/ and /v/ seemed decidedly *more voiced* than /ʒ/; indeed the latter could be heard as /ʃ/. The stimuli could not be altered to correct this feature

6.3.81
J. 710

without destroying the critical logic underlying the experiments in the first place, viz., using exactly the same stimuli for which Darwin (1971) had showed lateralized speech perception. However, if *S*s had been perceiving the voicing difference as we had, then the possibility existed that the recency and suffix effects were not based on fricative distinctions per se but rather on a voicing distinction.

To test this possibility the 22 stimuli from the entire set of 60 used in Experiment III were selected in which the last two positions included only /z/ and /v/ and the results were tabulated for this subset of the experimental data. Since the voicing distinction of which there is suspicion does not assist in making the /z/–/v/ discrimination it is important to know whether these stimuli yield results comparable to those of Figures 2 and 3. The answer is that they do: Correct recall probabilities in the control condition for *S*s who received fricatives alone combined with those who received fricatives accompanied by serial position were .23, .31, .37, .42, .44, .50, and .44 for the seven positions, respectively; and in the suffix condition, .20, .29, .34, .41, .44, .51, and .48, respectively. For the control condition the last-position recency effect was statistically significant, $T = 136.5$, $z = 1.97$, $p = .02$. In terms of absolute error frequencies, the suffix effect was not statistically significant on the last serial position, $T = 163.5$, $z = .90$. However, the error rates were lower absolutely in the early serial positions for the suffix condition than for the control condition. To correct for this, the proportion of all errors each *S* made on the last position was compared for the control and suffix conditions (see Crowder, 1971b); with this measure, the suffix caused a statistically significant reduction in relative recency, $T = 168.0$, $z = 1.90$, $p = .03$. Thus, restricting our attention to stimuli where no apparent voicing cue existed in the last two positions did not affect the pattern of results, though both the strength of the recency effect and of the suffix effect were slightly reduced numerically.

Although the recency and suffix effects found here with fricatives were statistically reliable they were small and many *S*s failed to show them. Data are not yet available to permit a statistical defense that these effects were smaller here than with other vocabularies, but a glance at any of the previous literature on suffix effects certainly suggests such an interaction. One strategy would be to sample different vocabularies of fricatives, including real-speech examples as well as synthetic ones in an effort to show that the more robust PA effects usually obtained can occur with fricatives as well. However, such an approac places the logical structure of the research i doubt for the reason mentioned above—th stimuli used have to be exactly the same one for which laterality effects are shown to occu in perception experiments. One obvious solution is a program of parallel research on perception and memory involving numerous different vocabularies. It is to be hoped that large experimental effects in such studies will suppor the inferences being drawn here.

Research in this area would be considerabl easier if special mechanisms of speech processing such as the right-ear advantage wer inevitably associated with stimuli and no affected by the task in which such stimuli ar used. Unfortunately this is not strictly th case with some types of stimuli. Haggar (1971) in particular has shown how vowe discriminations show varying degrees of laterality depending on the situational context. I has been shown (Pisoni, 1971), to cite anothe example, that simply shortening vowels tend to lead to increasingly categorical perception The fact that speech sounds can be made to "migrate" in and out of such special processing modes as the right-ear advantage suggests some caution in moving as far distant from perception experiments as are studies of immediat memory.

On the other hand, small as the effects ar in Experiments II and III and granted th huge difference between experimental contex in Darwin's (1971) laterality study and th present investigation, the results still wer statistically reliable and no feature of th present investigation is known to reduce th lateralization of speech sounds; therefore, th best course seems to be to accept tentativel the conclusion that the steady-state nature vowels led to their PAS representation rath than their being generally excluded from t types of special processing received ordinari by stop consonants.

General Discussion

The logic of the present research is aim at uncovering the reasons for different memo performance with different vocabularies speech sounds; it does not depend in any w upon which theory of the recency and suf effects turns out to be true. It is encouragi to the Crowder–Morton (1969) PAS theo however, to note that once again the presence and absence of recency and the presence and absence of the suffix effect are found to be perfectly correlated across vocabularies. In Table 2 the three main forms of data on which the PAS model rests (the modality, recency, and suffix effects) are displayed as a function of stimuli which have been investigated so far. Although not all cells of the matrix have as yet been filled with data, there are no exceptions to the rule that the putative symptoms of PAS appear together or not at all.

In terms of the PAS viewpoint it is tempting to conclude that PAS "holds" steady-state acoustic stimuli but not transient acoustic stimuli. As well as being premature in light of the small number of vocabularies so far investigated, this conclusion leads easily to the fallacious premise that transient speech sounds progress directly into a verbal short-term memory with no representation in a sensory memory. This must be false, for sensory storage is a logical necessity *precisely* for signals which are dynamic—otherwise, unless some representation of time t is still available at time $t + \Delta t$, the change itself could not be perceived. Sensory memory is thus a prerequisite for any cues depending on frequency analysis. However, the time orders involved in such storage are of a different order of magnitude than those postulated for PAS (Crowder, 1971b); cues for items such as stop consonants consist of sweeps lasting only a few dozen milliseconds. Whether PAS is continuous with such sensory storage, as is required for frequency analysis and for detection of transient signals, or whether it is a level removed from such fleeting storage is an issue remaining to be decided. It is relevant that other investigators have demonstrated sensory storage, with different properties from PAS, which is limited to the order of very short persistence required for these functions (e.g., Massaro, 1970). In any case, rather than conclude that PAS holds steady-state but not transient speech sounds, the more conservative conclusion is that it holds the former for longer time periods. This is not surprising—there are types of persistent afterimages in vision which depend upon the inspection figure being "burnt in" by relatively prolonged steady-state stimulation (Woodworth, 1938, p. 560).

TABLE 2

EVIDENCE FAVORING PRECATEGORICAL ACOUSTIC STORAGE (PAS) FOR DIFFERENT SPEECH VOCABULARIES

Stimulus vocabulary	Evidence of PAS		
	Auditory advantage over visual	Recency in auditory	Suffix effect at last position
Real speech			
Full alphanumeric[a]	+	+	+
Initial stop consonants[b]	−	−	no data
Terminal vowels[c]	+	+	no data
Synthetic speech			
Full alphanumeric[c]	no data	+	+
Initial stop consonants[b]	no data	−	−
Terminal stop consonants[d]	no data	−	−
Terminal vowels[c]	no data	+	+
Initial fricatives[d]	no data	+	+

Note. Symbols: (+) = data favoring PAS; (−) = data not favoring PAS.
[a] Crowder and Morton (1969).
[b] Crowder (1971a).
[c] Crowder (1971b).
[d] Present article.

REFERENCES

COLE, R. A. Different memory functions for consonants and vowels. *Cognitive Psychology*, 1973, **4**, 39–54.

CROWDER, R. G. The sounds of vowels and consonants in immediate memory. *Journal of Verbal Learning and Verbal Behavior*, 1971, **10**, 587–597. (a)

CROWDER, R. G. Waiting for the stimulus suffix: Decay, delay, rhythm, and readout in immediate memory. *Quarterly Journal of Experimental Psychology*, 1971, **23**, 324–340. (b)

CROWDER, R. G., & MORTON, J. Precategorical acoustic storage (PAS). *Perception & Psychophysics*, 1969, **5**, 365–373.

DARWIN, C. J. Ear differences in the recall of fricatives and vowels. *Quarterly Journal of Experimental Psychology*, 1971, **23**, 46–62.

FUJISAKI, H., & KAWASHIMA, T. Some experiments on speech perception and a model for the perceptual mechanism. *Annual Report of the Engineering Research Institute, Faculty of Engineering, University of Tokyo*, 1970, **29**.

HAGGARD, M. P. Encoding and the REA for speech signals. *Quarterly Journal of Experimental Psychology*, 1971, **23**, 34–45.

LIBERMAN, A. M., MATTINGLY, I. G., & TURVEY, M. T. Language codes and memory codes. In A. W. Melton & E. Martin (Eds.), *Coding processes in human memory*. New York: Halsted, 1972.

MASSARO, D. W. Perceptual processes and forgetting in memory tasks. *Psychological Review*, 1970, **77**, 557–567.

MATTINGLY, I. G. Experimental methods for speech synthesis by rule. *IEEE Transactions Audio and Electroacoustic*, 1968, **16**, 198–202.

MATTINGLY, I., LIBERMAN, A. M., SYRDAL, A. K., & HALWES, T. Discrimination in speech and nonspeech modes. *Cognitive Psychology*, 1971, **2**, 131–157.

MORTON, J., CROWDER, R. G., & PRUSSIN, H. A. Experiments with the stimulus suffix effect. *Journal of Experimental Psychology*, 1971, **91**, 169–190. (Monograph)

PISONI, D. B. On the nature of categorical perception of speech sounds. In, *Supplement to status report on speech research*. New Haven, Conn.: Haskins Laboratories, 1971.

SHANKWEILER, D. P., & STUDDERT-KENNEDY, M. Hemispheric specialization for speech perception. *Journal of the Acoustical Society of America*, 1970, **48**, 579–594.

SMALLWOOD, R. A., & TROMATER, L. J. Acoustic interference with redundant elements. *Psychonomic Science*, 1971, **22**, 354–356.

WOODWORTH, R. S. *Experimental psychology*. New York: Holt, 1938.

(Received May 1, 1972)

Journal of Experimental Psychology
1973, Vol. 98, No. 1, 25–30

ABSTRACT AND CONCRETE PHRASES IN FALSE RECOGNITION [1]

CONNIE GOLDFARB, JOYCE WIRTZ, AND MOSHE ANISFELD [2]

Yeshiva University

Adjective-noun phrases (termed preceding [P] phrases), followed by phrases bearing synonymous (S) and antonymous (A) relations to them, were presented auditorily to *S*s. They evaluated each phrase as to whether it had appeared before on the list or was new. All the P, S, and A phrases were new so that all errors entailed false recognition. More errors were made to the S phrases than to the A phrases. Since this result was obtained for both concrete and abstract material, it is taken to mean that verbal messages in general, not just image-evoking messages, are coded primarily by reference to their content rather than linguistic form. The $S > A$ effect was, however, weaker for abstract phrases than for concrete phrases. This was attributed to the relative indistinctness of abstract phrases. An independent experiment has demonstrated a strong $S > A$ effect for relatively distinct *abstract* phrases.

It has been suggested recently that abstract verbal material is coded differently in memory than concrete material (e.g., Bower, 1970; Paivio, 1969). According to the two coding systems hypothesis, concrete material is transformed into an imagery code, while abstract material is retained in a code closer to the verbal form in which it was presented. The present research provides evidence for an alternative hypothesis which claims that all verbal material is coded similarly, for referential meaning, and attributes differences between concrete and abstract material to differences in distinctness rather than to differences in imagibility.

The two coding systems hypothesis can be exemplified by reference to an experiment by Begg and Paivio (1969) which compared the recognition memory of abstract and concrete sentences. After initial presentation of a set of sentences, semantic or lexical changes were introduced in some of them and *S*s were required to indicate for each sentence whether it was identical or changed. They found that the semantic changes were more readily noticed in the concrete sentences than in the abstract sentences, and that there was an opposite trend for the lexical changes. Begg and Paivio interpreted the results in terms of different modes of coding for abstract and concrete material. Concrete sentences, they argued, were transformed into an imaginal code consisting of a unitized image for each sentence, whereas abstract sentences were retained in terms of the individual words composing them. Bower (1970, p. 508) offered a related interpretation for Gorman's (1961) finding that concrete nouns were more accurately recognized than abstract nouns.

An alternative to the two coding systems hypothesis to account for differences between abstract and concrete material in recognition memory arises from a study by Anisfeld (1970). That experiment used a false recognition technique for inferring underlying memorial representation. In this technique, *S*s are presented with a continuous list of items and for each item they have to indicate whether it appeared before on the list or was new. A false recognition error is one involving an old response to a new item. It was found that *S*s made more false recognition errors to adjective-noun phrases (e.g., REAR DOOR) synonymous with previously presented phrases (BACK DOOR) than to antonymous phrases (FRONT DOOR). In fact, the number of errors to the antonymous phrases did not differ from the number of errors to unrelated control phrases. This finding makes sense if one assumes that the

---

[1] This research was supported by National Science Foundation Grant GB-21280, and is an expanded version of a paper presented at the meeting of the Eastern Psychological Association, Boston, April 1972.

[2] Requests for reprints should be sent to Moshe Anisfeld, Department of Psychology, Yeshiva University, New York, New York 10003.

TABLE 1

MEAN IMAGERY AND FREQUENCY RATINGS AND FALSE RECOGNITION SCORES FOR PHRASES IN EXPERIMENT I

Phrase	Rating		False recognition score
	Imagery	Frequency	
Abstract			
P	2.21	6.57	.39
S	2.26	6.62	1.29
A	2.31	6.61	1.02
Concrete			
P	4.26	6.47	.23
S	4.27	6.49	1.49
A	4.22	6.54	.82

Note. Each set of phrases contained a preceding phrase (P), a synonymous phrase (S), and an antonymous phrase (A). All means are per *S* per phrase.

phrases were not retained in the verbal form in which they were presented but rather were recoded in terms of their meaning. The referents of synonymous phrases are more similar, and hence more likely to be confused with one another, than the referents of antonymous phrases. The finding of no false recognition effect for antonymous *phrases* contrasted with the effect obtained for antonymous *words* in the same study as well as in other studies. It was proposed (Anisfeld, 1970) that the referential mode of coding was more dominant for phrases than for words because phrases have more definite and determinate referents than words.

The interpretation offered to explain the difference in recognition memory between words and phrases can be extended to abstract and concrete items, because abstract items tend to have less definite denotations than concrete items (e.g., Martin, 1969). Their indefinite denotations render abstract words less distinct than concrete words.

There is another factor that contributes to the relative indistinctness of abstract words. Abstract words tend to label important concepts which are deeply embedded in the individual's cognitive and emotive systems and closely interrelated with other concepts. One reflection of this involvement and interrelatedness was reported by Paivio and Begg (1971). They found that the average score for abstract words on a scale of associative interrelatedness was almost double the score for concrete words. However, involvement and relatedness are not inseparable qualities of abstractness, and many abstract words are no more conceptually involved than concrete words.

In addition to its difficulty in evoking images, abstract material is thus characterized by two other qualities, which can be viewed as two aspects of relative indistinctness: indefiniteness of denotation and higher interrelatedness with other concepts. The difference in recognition-memory results between abstract and concrete material has been attributed to the difference in imagery. But this difference can be more parsimoniously accounted for in terms of the distinctness factor, since this factor would also account for differences in recognition memory between frequent and infrequent words (e.g., Schulman, 1967). Distinctness has, in fact, been mentioned by several authors as a possible explanation for the frequency effect (e.g., Paivio, 1969; Schulman, 1967). Wickens and Engle (1970) have introduced evidence in favor of the distinctness hypothesis in the Peterson and Peterson (1959) STM task.

The present research attempted to test the distinctness hypothesis in the case of recognition memory. The first experiment simply compared the false recognition of a group of undifferentiated abstract phrases with a group of concrete phrases, using preceding, synonymous, and antonymous phrases as in the Anisfeld (1970) experiment described above. The second experiment compared the false recognition of relatively indistinct abstract phrases with the false recognition of relatively distinct abstract phrases. If differences between abstract and concrete material in recognition memory are due to distinctness, as is hypothesized here, then the results for the indistinct abstract phrases in the second experiment should duplicate the results obtained in the first experiment for the undifferentiated (typical) abstract phrases whereas the results for the distinct abstract phrases should be more like the results for the concrete phrases in the first experiment.

In summary, the present research is dedicated to the proposition that all verbal messages, not just concrete messages, are coded primarily for their meaning. The empirical index used here for semantic coding is a larger number of false recognition errors to phrases synonymous with previously heard phrases than to antonymous phrases. This research further attempts to show that to the extent that there are differences in recognition memory between abstract and concrete phrases they are due to differences in denotative distinctness rather than to differences in imagibility.

EXPERIMENT I

Method

Subjects. In all, 78 *Ss* were used in the preparation of material and in the experiment proper. About half of the *Ss* were male and half were female. They were paid volunteers who met the requirements of being under 30, having had some college education, and being native English speakers. Twenty *Ss* rated the phrases for imagery, 20 *Ss* rated them on relative frequency of occurrence, 4 *Ss* checked the auditory intelligibility of the material, and 34 *Ss* (22 male, 12 female) took part, individually, in the experiment proper.

Materials and procedure. Twenty-four sets of phrases were constructed for the abstract category and 24 for the concrete category. Each set contained a preceding phrase (P), a synonymous phrase (S), and an antonymous phrase (A). Care was taken that there would be no overlap between the concepts represented by the various sets of phrases.

This initial pool of phrases was randomized and presented to 20 *Ss* in written form for rating on an imagery-evokedness scale. The procedure for obtaining imagery values was similar to that employed by Paivio, Yuille, and Madigan (1968), except that a 5-point scale was used instead of their 7-point scale. Imagery ratings were used to classify the phrases as abstract and concrete because these ratings largely overlap with concreteness ratings and tend to have better predictive power (Paivio, 1968).

Another group of 20 *Ss* rated the phrases, again in written form, on a 10-point scale of relative frequency of occurrence. The scale ran from 1 (Very Infrequent) to 10 (Very Frequent). The *Ss* were instructed to base their ratings on how frequently they felt the phrases appeared in everyday writing and conversation.

With the imagery and frequency information available, 10 abstract and 10 concrete sets of phrases were selected for the experiment. Each set included one P, one S, and one A phrase. Table 1 presents the mean imagery and frequency values for the abstract and concrete P, S, and A phrases. As can be seen in the table, the three subgroups within the abstract category and within the concrete category were closely matched on imagery and all six subgroups were matched on frequency. A sample abstract set is: IMMEDIATE RESPONSE (P), PROMPT RESPONSE (S), DELAYED RESPONSE (A). A sample concrete set is: FILTHY SHIRT (P), DIRTY SHIRT (S), CLEAN SHIRT (A). The intelligibility of the phrases was checked by recording them on tape and having four *Ss* transcribe them. All of the phrases used in the experiment were accurately reported.

In addition to the experimental phrases described, 48 filler phrases, 24 abstract and 24 concrete, were also used. Within each category, 4 filler phrases appeared once, 10 twice, and 10 three times. There was a total of 148 phrases. They were recorded on magnetic tape at the rate of 1 phrase every 5 sec.

The first 20 positions on the tape were occupied by filler phrases. Otherwise, the abstract and concrete P phrases were spread randomly throughout the list. Each P phrase was followed within 20–40 positions by a related phrase (S or A), with a mean separation of 31.4 positions. Two different test lists were prepared so that each *S* would receive all P phrases, but only one related phrase from each set. Each *S* received 10 S phrases (5 abstract and 5 concrete) and 10 A phrases (5 abstract and 5 concrete). The positions occupied by A phrases in one list were occupied by S phrases in the other list, and vice versa.

The *Ss* were told that they would hear a list of two-word phrases and that for each phrase they were to decide whether or not they had heard it earlier on the list. They were to rate as 1 phrases they judged to be "definitely new," as 2 phrases that they thought were "probably new," as 3 phrases that were "probably old," and as 4 phrases they judged to be "definitely old."

Results and Discussion

The responses were scored in terms of the number of points by which the number given by *S* to the P, S, and A phrases deviated from 1, the correct response for all of these phrases. The mean deviation scores for the P, S, and A phrases can be seen in Table 1. Since they were not preceded by any specifically built-in relations, the error rate to the P phrases was quite low, but it was higher for the abstract category than for the concrete category, $t\ (33) = 12.1$, $p < .001$. To the extent that errors are made to the P phrases, they must be due to relations with other phrases in the list or with material outside the list, i.e., to relations not manipulated in the experiment. Being less definite, the abstract phrases are more likely to remind *S*s of experimentally unspecified concepts while the concrete phrases, having a clearer identity and closure, are less likely to trigger other extraexperimental concepts.

To compare the false recognition values for S and A concrete phrases with those of S and A abstract phrases, a 2 × 2 × 2 analysis of variance (List × Abstractness–Concreteness × Synonymy–Antonymy) was carried out. It showed no significant effect for list or abstractness–concreteness, but the main effect for synonymy–antonymy was clearly significant, $F\ (1,\ 32) = 43.6,\ p < .01$. That is, the error rate for the S phrases was higher than for the A phrases. None of the interactions was significant. But a test for simple main effects showed that the difference between S and A phrases within the concrete category was highly significant, $F\ (1,\ 32) = 12.5,\ p < .01$, whereas the difference within the abstract category was not, $F\ (1,\ 32) = 2.0$, although it was in the same direction.

The main effect showing a higher error rate for the S than for the A phrases, and its failure to interact with abstractness-concreteness, indicates that both the abstract and concrete phrases were coded similarly, mainly in terms of their semantic content. The weakness of the S > A effect for the abstract category may be attributed to the greater role played in this category by extraexperimental relations (as indicated by the higher number of errors for the abstract P phrases than for the concrete P phrases), leading to a reduction of the influence of the P relation on the abstract S and A phrases. In other words, the abstract S and A phrases, because of their greater indistinctness and diffuseness, are assumed to have been less tied than the concrete S and A phrases to their P mates and more tied to other concepts in *S*'s background. Since the difference in results between S and A is predicated on their different relations to P, any diminution of the influence of P would reduce the difference between them. If this interpretation is valid, then a reduction of the extraexperimental error component should enhance the S > A effect. Since the larger extraexperimental component for abstract phrases was attributed to their relative indistinctness, an attempt was made in the second experiment to raise the distinctness of abstract phrases.

Distinctness of phrases in the second experiment was manipulated via a variable of personal relevancy. Two groups of phrases were selected: One group was rated more personally relevant (MR) and the other less relevant or of little personal significance (LR). This variable is based on the idea that in memorial coding new messages are assimilated into and integrated with the individual's personal schemata (Bartlett, 1932). On this assumption, one might expect that the MR phrases will be readily assimilated into *S*s' personal schemata and later, in a memory test, it will be relatively difficult to disentangle them from the conceptual framework into which they were assimilated. The LR phrases, on the other hand, are not expected to find a ready place in *S*s' personal schemata and hence are more likely to be retained in a form more directly reflecting their uniquely distinct characteristics.

EXPERIMENT II

Method

Subjects. In all, 114 *S*s, about half male and half female, were used in this experiment. They met the same general qualifications as *S*s in the previous experiment. Of the 114, 40 *S*s rated the phrases for personal relevancy, 40 rated them for imagery, and 34 (17 male, 17 female) took part in the experiment proper. The *S*s who rated the phrases were tested in large groups and the *S*s who participated in the experiment proper were tested in groups of 1–3 and paid for their services.

Materials and procedure. An initial pool of 94 sets of phrases was arbitrarily divided in half. Each of the two halves was recorded on tape at the rate of 5 sec. per phrase. Two other tapes, using the same words in reverse order, were also prepared. Four tapes were thus used for the ratings. Each phrase on the tapes was rated by 20 *S*s on imagery and by 20 different *S*s on relevancy. For the imagery ratings, the same 5-point scale and instructions were used as in the first experiment. The relevancy scale also had 5 points: 1 denoting very low relevancy and 5 very high relevancy. In the instructions to the relevancy scale *S*s were asked to assign a number to each phrase on the basis of the extent to which the idea had a personal meaning or reference for them.

In view of the personal character of the relevancy variable, it was important to check whether the relevancy ratings obtained from one group of Ss were applicable to another group. For this purpose, the mean relevancy values given each of the 282 phrases (94 sets, with 3 phrases in each set) by half of the *S*s (10) were correlated with those given them by the other half. A respectable split-half reliability coefficient was obtained, r (281) = .64, $p < .001$, justifying the employment of the relevancy ratings in the construction of materials for use in the false recognition experiment.

From the 94 sets, 20 sets of abstract phrases were chosen: 10 more relevant (MR) and 10 less relevant (LR). The mean relevancy and imagery values for the P, S, and A MR and LR phrases are presented in Table 2. As can be seen in the table, the P, S, and A phrases within the MR and within the LR categories were closely matched on relevancy, and all six subgroups were equated on imagery. A sample MR set is: PERSONAL APPROACH (P), SUBJECTIVE APPROACH (S), OBJECTIVE APPROACH (A). A sample LR set is: ARRESTED DEVELOPMENT (P), RETARDED DEVELOPMENT (S), PRECOCIOUS DEVELOPMENT (A).

Two tapes were prepared for the experiment. Each tape consisted of all 20 P phrases, 5 LR S phrases, 5 LR A phrases, 5 MR S phrases, and 5 MR A phrases. The 5 S and 5 A phrases for the MR and LR categories were randomly selected for the first tape, with the restriction that no 2 phrases belong to the same set. Tape 2 was identical to Tape 1 except that where Tape 1 had an S phrase, Tape 2 had its corresponding A phrase, and where Tape 1 had an A phrase, Tape 2 had its corresponding S phrase.

Forty-four filler phrases, 22 MR and 22 LR, were selected from the original pool, with care taken that there be no relationship between any of the fillers and the experimental (P, S, or A) phrases. To establish a set for repetitiveness, 11 of the MR fillers and 11 of the LR fillers were presented twice, and the other 11 in each category were presented three times. Thus, each tape consisted of 150 phrases: 40 experimental phrases (20 P phrases, 10 S phrases, and 10 A phrases) and 110 filler phrases. The phrases were recorded at the rate of 1 phrase every 5 sec.

The first 20 positions on each tape were occupied by filler phrases. The P phrases were randomly distributed throughout the rest of the list. The S or A phrase related to each P phrase was randomly assigned a position 20–40 spaces away. The mean separation was 31.2 spaces.

Half of the *S*s (17) received Tape 1 and half Tape 2. The *S*s were instructed to rate each phrase as either O (old: appeared before on the tape) or N (new: did not appear before on the tape). If uncertain, they were to give the most probable answer. This experiment thus used a dichotomous scale in contrast to the first experiment which employed a 4-point scale. It has been estimated that the 4-point scale is only marginally more sensitive than the dichotomous scale (Anisfeld, 1970).

TABLE 2

MEAN RELEVANCY AND IMAGERY RATINGS AND FALSE RECOGNITION SCORES FOR PHRASES IN EXPERIMENT II

Phrase	Rating		False recognition score
	Relevancy	Imagery	
More relevant			
P	3.58	2.84	.13
S	3.56	2.85	.41
A	3.43	2.81	.32
Less relevant			
P	2.50	2.85	.04
S	2.46	2.87	.36
A	2.43	2.84	.19

Note. Each set of phrases contained a preceding phrase (P), a synonymous phrase (S), and an antonymous phrase (A). All means are per *S* per phrase.

Results and Discussion

For all experimental phrases (P, S, and A) the correct response was N. Each O response to these items constituted one false recognition error. Table 2 shows the mean false recognition values for the six subcategories of phrases. Statistical comparison shows that the error rate to the P phrases was significantly lower in the LR category than in the MR category, t (33) = 3.1, $p < .01$. The minimal error rate to the LR P phrases reflects, according to the present interpretation, the low influence of extraexperimental relations on the distinct phrases of this category.

A 2 × 2 × 2 analysis of variance (List × Relevancy × Synonymy–Antonymy) shows no significant difference between the two tapes, a significantly higher error rate for the S phrases than for the A phrases, F (1, 32) = 42.8, $p < .01$, and for the MR phrases than for the LR phrases, F (1, 32) = 6.6, $p < .05$. There was no significant interaction between any two of the three factors. A test for simple main effects shows that the difference between S and A phrases was highly significant for the LR category, F (1, 32) = 13.8, $p < .01$, but was not significant for the MR category, F (1, 32) = 3.5, although it was in the same direction.

This pattern of results for the comparison between abstract MR and LR phrases shows a parallel with the results of the first experiment comparing undifferentiated abstract phrases with concrete phrases. The MR category behaved like the abstract category and the LR category behaved like the concrete category. More false recognition errors were made to MR and abstract P phrases than to LR and concrete P phrases, respectively; and the S > A effect was weaker for the MR and abstract phrases than for the LR and concrete phrases, respectively. MR phrases are less distinct than LR phrases in the sense that they are more likely to become assimilated into *S*s' overall cognitive structures and thus lose, in part, their individual identities. Abstract phrases are more indistinct than concrete phrases because their denotations are less clearly defined and circumscribed. In both experiments, the less distinct phrases attracted more experimentally unspecified sources of error, reducing the S > A effect. Also, to the extent that indistinct P phrases do not retain their specific referential identities in memory, their "pull" on the S and A phrases is likely to diminish. The argument thus is that the indistinct S and A phrases, in comparison to the distinct phrases, were more influenced by extraexperimental relations and less by the P relations.

CONCLUSION

In conclusion, the present research has indicated that lexical coding is no more indigenous to abstract material than to concrete material; both are coded primarily for referential meaning. Differences in recognition memory between concrete and abstract items can be more parsimoniously attributed to differences in distinctness than to differences in imagibility. It was proposed that distinct items in general (not just concrete items) attracted fewer false recognition errors because they did not become fully assimilated into *S*'s cognitive schemata and retained their original identities. In contrast, indistinct items became more fully integrated into *S*'s schemata and were therefore more readily confused with other items in the system.

REFERENCES

ANISFELD, M. False recognition of adjective-noun phrases. *Journal of Experimental Psychology*, 1970, **86**, 120–122.

BARTLETT, F. C. *Remembering*. Cambridge, England: Cambridge University Press, 1932.

BEGG, I., & PAIVIO, A. Concreteness and imagery in sentence meaning. *Journal of Verbal Learning and Verbal Behavior*, 1969, **8**, 821–827.

BOWER, G. H. Analysis of a mnemonic device. *American Scientist*, 1970, **58**, 496–510.

GORMAN, A. M. Recognition memory for nouns as a function of abstractness and frequency. *Journal of Experimental Psychology*, 1961, **61**, 23–29.

MARTIN, J. E. Semantic determinants of preferred adjective order. *Journal of Verbal Learning and Verbal Behavior*, 1969, **8**, 697–704.

PAIVIO, A. A factor-analytic study of word attributes and verbal learning. *Journal of Verbal Learning and Verbal Behavior*, 1968, **7**, 41–49.

PAIVIO, A. Mental imagery in associative learning and memory. *Psychological Review*, 1969, **76**, 241–263.

PAIVIO, A., & BEGG, I. Imagery and associative overlap in short-term memory. *Journal of Experimental Psychology*, 1971, **89**, 40–45.

PAIVIO, A., YUILLE, J. C., & MADIGAN, S. Concreteness, imagery, and meaningfulness values for 925 nouns. *Journal of Experimental Psychology Monograph Supplement*, 1968, **76** (1, Pt. 2).

PETERSON, L. R., & PETERSON, M. J. Short-term retention of individual verbal items. *Journal of Experimental Psychology*, 1959, **58**, 193–198.

SCHULMAN, A. I. Word length and rarity in recognition memory. *Psychonomic Science*, 1967, **9**, 211–212.

WICKENS, D. D., & ENGLE, R. W. Imagery and abstractness in short-term memory. *Journal of Experimental Psychology*, 1970, **84**, 268–272.

(Received May 19, 1972)

Journal of Experimental Psychology
1973, Vol. 98, No. 1, 31–35

ANALYSIS OF LETTER STRINGS IN WORD RECOGNITION[1]

ROBERT F. STANNERS[2] AND GARY B. FORBACH

Oklahoma State University

The major concerns of the present experiment were to determine whether or not letter strings are analyzed into components in the course of making a word–nonword decision, and, if so, what role the components have in addressing memory. Three categories of material were used: consonant-consonant-vowel-consonant-consonant nonwords (CCVCCs), consonant-consonant-vowel-consonant-consonant words (WORDs), and consonant-consonant-consonant-consonant-consonant items (CCCCCs). Initial and terminal consonant pairs were matched across the three categories of material, e.g., CRUSS, CROSS, and CRNSS. Within categories the frequency of occurrence in the language of the initial and terminal consonant pairs was varied. The frequency of occurrence of the nonwords as a unit was always zero. An analysis hypothesis was supported in that within both the categories CCVCC and CCCCC decision latency was directly related to component frequency. Across the categories the shortest overall latency was for CCCCCs, followed by WORDs and then CCVCCs. The interpretation of these and other results was that information from the consonant pairs is used to address a subset of memory representations and that the size of the subset is related to the frequency of occurrence of the consonant pairs.

A recent experiment by Stanners, Forbach, and Headley (1971) was concerned with the type of information in a letter string which affects the time required either to accept or reject it as a vocabulary member. The materials were visually presented three-letter items which were consonant-vowel-consonant words, phonologically lawful nonwords (CVCs), or three-letter consonant strings (CCCs). The *Ss* were instructed to indicate as quickly as they could by means of a switch movement whether an item was in their vocabulary or not. The results were compatible with a type of recognition process in which the item is phonologically edited and may be rejected at this stage of processing. If the item is found to be phonologically lawful, a search process ensues through a subset of memory the size of which is determined by the frequency of occurrence in the language of the initial and terminal consonants. However, a check of the frequency of the CVCs as complete units in the norms of Mayzner and Tresselt (1965) gave rise to another possible interpretation. The frequency of the CVCs as units, i.e., as portions of words, was found to be positively related to classification latency. Thus, there is a possibility that the individual consonants were not used to index a subset of memory representations but that the item was first encoded as a unit, and the size of the memory subset is related to unit frequency rather than to the frequency of the components of the unit.

A principal concern of the present experiment was to determine whether or not an item is analyzed into components during the classification task. If analysis does take place, then certain linguistic properties of the components, such as frequency of occurrence in the language, might affect the addressing of the appropriate memory representation or subset of representations. A straightforward means of testing an analysis hypothesis would be to remove the covariation between unit frequency and component frequency. Five-letter nonwords of the form consonant-consonant-vowel-consonant-consonant (CCVCC) have a unit frequency of essentially zero in the language. However, the frequency of components, initial and terminal consonant pairs, can be varied over a wide range. If

---

[1] This research was partially supported by a grant from the Oklahoma State University Research Foundation.

[2] Requests for reprints should be sent to Robert F. Stanners, Department of Psychology, Oklahoma State University, Stillwater, Oklahoma 74074.

the frequency of initial and terminal consonant pairs in CCVCCs had an effect analogous to that of the single consonants of the Stanners et al. (1971) study, it would seem that *S* must analyze the letter string to have his latency affected by the frequency of the components. Furthermore, the use of five-letter consonant strings (CCCCCs) in which the frequency of initial and terminal consonant pairs was varied would provide another strong test of an analysis hypothesis. The analysis hypothesis would be upheld if the frequency of initial and terminal consonant pairs affected classification latency since a CCCCC item certainly has a unit frequency of zero in the language and does not even appear to be encodable as a unit, at least in the very short times involved in the classification task.

METHOD

Subjects. Twenty undergraduates at Oklahoma State University served as *S*s. The *S*s were given a small bonus toward their final grade in introductory psychology as an inducement to participate.

Materials. Three categories of materials were used, consonant-consonant-vowel-consonant-consonant nonwords (CCVCCs), consonant-consonant-vowel-consonant-consonant words (WORDs), and consonant-consonant-consonant-consonant-consonant items (CCCCCs). The reference for the construction of the materials was a set of norms compiled by Venezky (1962) which is a frequency analysis of the spelling-to-sound correspondence of the English language. The frequency with which a letter or letter combination occurred as a phoneme or phoneme combination in a given position was tabulated for a sample of approximately 20,000 words from the *Thorndike Century Senior Dictionary* (1941). A somewhat more detailed description of the norms may be found in Stanners (1970). For the purposes of the present study the data of interest were the frequencies of initial and terminal consonant pairs based on their occurrence as specific sound units in words.

Sets of items composed of a WORD, a CCVCC, and a CCCCC were constructed which had identical initial and terminal consonant pairs, e.g., CROSS, CRUSS, and CRNSS. The sets were composed in one of four different ways, with a relatively high-frequency initial consonant pair and high-frequency terminal consonant pair (HH), with a high-frequency initial consonant pair and low-frequency terminal consonant pair (HL), and, in similar fashion, LH and LL. There were 15 exemplars of CCVCCs, WORDs, and CCCCCs for each of the four frequency combinations with the exception of LL WORDs which are very rare; only two of these could be constructed. Thirteen filler WORDs were selected to make up the deficit in the LL group as well as an additional 60 filler WORDs to equalize the number of words and nonwords in the total sample of 240 items.

The average frequency in the Venezky (1962) norms of the initial consonant pair for the HH items was 201.7 and the range was 95–468. The counterparts of these statistics for the HL, LH, and LL items were, respectively, 165.0 and 79–468, 28.7 and 6–39, 6.3 and 1–21. For terminal consonant pairs the frequency statistics were as follows for HH, HL, LH, and LL: 289.6 and 97–729, 10.8 and 1–17, 216.2 and 97–376, 3.3 and 1–8. The Mayzner and Tresselt (1965) norms were used to check the frequency of occurrence of the CCVCCs and CCCCCs as units. All of the CCVCCs except two had a frequency of occurrence of zero; the two exceptions had a frequency of one. All CCCCCs had frequencies of occurrence of zero.

The materials were typed in uppercase with an IBM sign typewriter, reproduced onto transparencies by the diazochrome method, and mounted in 35-mm. slide holders.

Procedure and design. The *S* was seated at a small table in front of a Plexiglas screen onto which the materials were back-projected to produce a visual angle of approximately 4°. The basic programming and recording equipment consisted of two Kodak Carousel projectors equipped with solenoid-operated shutters and controlled by an eight-channel Lafayette timer. A Hunter Klockounter was used to measure the latencies.

A trial was begun by *S*'s pressing a thumb switch held in his nonpreferred hand. The *S* had been instructed that before he pressed the thumb switch he should attend closely to the screen and hold between the thumb and forefinger of his preferred hand a lightly sprung toggle-type switch. Following activation of the thumb switch by 1 sec., an item was presented on the screen. The *S* was instructed to indicate by the direction (left or right) of the switch movement whether the item as a complete unit was or was not in his vocabulary. The direction of movement was indicated on a sign next to the switch and was held constant for a given *S*. The Klockounter started with the presentation of the item and stopped with *S*'s response. Both speed and accuracy were stressed by the instructions, and responding positively to idiosyncratic nicknames was discouraged. The offset of a lamp 3 sec. after *S*'s response signaled that he could begin a new trial whenever he was ready.

An experimental session for a given *S* consisted of 280 trials (including 40 practice trials) and lasted approximately 55 min. The items were randomly ordered for each *S*. The experimental design was completely within *S*s, and the variables were frequency of the initial letter (I), frequency of the terminal letter (T), and category (C). Direction of switch movement for signification of whether an item was or was not in *S*'s vocabulary was balanced between *S*s.

TABLE 1

Average Latency of Correct Responses by Subcondition and Average Latency of Correct Responses and Errors by Category

Category	Frequency of consonant pair				Overall average	
	HH	HL	LH	LL	Correct responses	Errors
CCVCC	866	799	831	653	788	718
WORD	641	723	737		700	799
CCCCC	646	628	643	586	626	

Note. Empty cells are those for which the mean is based on too few cases.

Results

A logarithmic transformation was performed on the data to reduce the effect of one or a few highly deviant latencies on an *S*'s score. An individual score in the major data analyses was the antilogarithm of the mean of the log latencies for a given *S* in a subcondition. Only "correct" (in agreement with the category system) responses were used in the computation of these scores. The error rate was such that all scores were based on at least 10 latencies (out of a possible 15) and 97% of them were based on 12 or more latencies.

The means for all subconditions are presented in Table 1. The first type of analysis was a three-factor analysis of variance involving CCVCCs and CCCCCs. WORDs were not included because of the empty cell for LL. In this analysis, as well as in all others, only effects with probability values beyond the .05 level were considered significant. All main effects were highly significant; for I, $F\ (1, 19) = 34.97$, $p < .001$; for T, $F\ (1, 19) = 64.47$, $p < .001$; and for C, $F\ (1, 19) = 82.37$, $p < .001$. All interaction effects were significant as well. The two-factor terms were I × C, $F\ (1, 19) = 34.00$, $p < .001$; T × C, $F\ (1, 19) = 17.08$, $p < .001$; and T × I, $F\ (1, 19) = 24.52$, $p < .001$. The significant three-factor interaction term indicated that even though the patterning of the means was very similar for the CCVCCs and the CCCCCs, the differences among the means were much larger for the CCVCCs; for C × I × T, $F\ (1, 19) = 6.13$, $p < .025$. Subanalyses were conducted to deal with the effects within CCVCCs and CCCCCs.

Within the category of CCVCCs all effects were significant. The main effects were I, $F\ (1, 19) = 40.84$, $p < .001$; and T, $F\ (1, 19) = 40.68$, $p < .001$. The larger difference between the LL and LH means as compared to the HL and HH means is reflected in a significant interaction term for I × T, $F\ (1, 19) = 16.26$, $p < .001$.

All effects within the category of CCCCCs were likewise significant, reflecting the basically similar patterning of subgroup means for the two categories of items; for I, $F\ (1, 19) = 10.09$, $p < .005$; and for T, $F\ (1, 19) = 36.33$, $p < .001$. The significant interaction term (I × T) reflected the same general pattern as for the CCVCCs, $F\ (1, 19) = 13.52$, $p < .005$.

The Newman-Keuls procedure (Kirk, 1968, pp. 91–93) was used to test differences between subgroup means within the category of WORDs and to make some tests between WORDs on the one hand and CCVCCs and CCCCCs on the other. Within WORDs, the means for LH and HL were both significantly different ($p < .05$) from HH but not from each other. The comparisons between categories involved testing the three means for WORDs against their counterparts for CCVCCs and CCCCCs; e.g., *HH* WORDs vs. *HH* CCVCCs and *HH* CCCCCs, *LH* WORDs vs. *LH* CCVCCs and *LH* CCCCCs, etc. All the differences were significant beyond the .01 level except for that between HH WORDs and HH CCCCCs which was not significant.

Responses were designated errors if they failed to agree with the category system. The error rate for WORDs was .067; for

CCVCCs, .060; and for CCCCCs, .004. The number of errors per *S* was too small for statistical analysis of the latencies, but the mean latencies collapsed over subconditions and *S*s were calculated for the CCVCCs and WORDs and are presented in Table 1.

DISCUSSION

The results indicate a number of the properties required by a memory indexing model to account for word–nonword classification. The results for CCCCCs strongly indicate that analysis of the item does take place during classification. Differences in latency are related to differences in consonant pair frequency and are statistically highly reliable. At the same time, unit frequency must be the same for the CCCCCs, i.e., zero. The very large overall difference in latency between CCCCCs and CCVCCs raises a question concerning the role of the consonant pairs in the classification task for the two types of items. If a search process is taken to mean the actual comparison of a stimulus item with a subset of memory representations, then it does not appear that the frequency of the consonant pairs affects the search process in the case of CCCCCs. Rubenstein, Lewis, and Rubenstein (1971) have presented evidence indicating that the comparison of the item with the memory representation involves a phonologically recoded version of the item. Nonwords which were homophonic with words, e.g., BRANE, WATE, were found to have longer rejection latencies than comparable nonhomophonic nonwords. Presumably, after finding a phonological match, *S* had to reject the item on the basis of orthographic differences. In the same study it was found that homophonic words, e.g., ATE-EIGHT, had longer acceptance latencies than nonhomophones. The interpretation was that the extra time was spent in sometimes finding a phonological match which was not an orthographic match.

Since the CCCCCs are not phonologically encodable, the effects attributable to the frequency of the consonant pairs must reflect a process other than comparison. A possibility suggested by Rubenstein et al. (1971) is "marking," which involves the designation of the memory subset to be searched. Thus, the higher the frequency of the consonant pair, the larger the subset to be marked and the greater the amount of time involved. A relationship between subset size and frequency would be compatible with the type of frequency count on which the Venezky (1962) norms are based. In contrast to most letter counts, the Venezky norms did not use running text as its source but a dictionary. Therefore, the frequencies are based on the number of *different* words in which a given letter or letter combination occurs. Presumably, the number of different words in which a consonant pair occurred would be positively related to the size of the memory subset which would have to be marked. Processing of a CCCCC would involve marking a subset of memory and then attempting a phonological encoding of the item. Rejection of an item could take place at the encoding stage, and this early rejection would be the basis for the much lower overall latencies for CCCCCs as compared to CCVCCs. Rejection of CCCCCs prior to search is also supported by error data; about 16 times as many errors were made on CCVCCs and WORDs as compared to CCCCCs.

The frequency effects observed for the CCCCCs in the present experiment did not occur in the Stanners et al. (1971) study for the case of CCCs. Although there is no obvious reason for the discrepancy, one possibility is that the absolute frequency of individual initial and terminal consonants is fairly high even in the low-frequency subconditions. Where the overall average frequency of initial low-frequency consonant pairs in the present study is 17.5, that of initial low-frequency single consonants in the earlier study is 346.0. The counterparts of these values for low-frequency terminal units are 7.0 and 108.0.

The results for the CCVCCs as well as those for the CCCCCs support an analysis hypothesis since there are large effects attributable to the frequency of the consonant pairs while unit frequency for CCVCCs is essentially zero. Since CCVCCs are phonologically encodable they are involved in the comparison or search process through the marked subset, and the latencies would reflect marking time plus search time. Frequency of the consonant pairs in determining the size of the memory subset would affect both marking and search time, although the results for the CCVCCs indicate that the search process is a much slower one than marking. Rejection of a CCVCC would require an exhaustive search through the appropriate subset of memory representations.

The patterning of the means for the CCCCCs and CCVCCs is similar and produced a significant I × T interaction in each case (se

Results). For both CCCCCs and CCVCCs he pattern would be interpreted as a reflection f the relationship between consonant pair frequency and size of the memory subset for he present materials. Since the duration of marking and search are both dependent on memory subset size, the pattern would be expected to be similar for CCCCCs and CCVCCs.

The search through the marked subset is not exhaustive for WORDs but terminates with the detection of a match. A nonexhaustive search process for WORDs is compatible with the finding that the mean latency for a given frequency combination for WORDs falls somewhere between its counterparts for CCCCCs and CCVCCs. A nonexhaustive search process for WORDs is also supported by error data. The average latency of the errors for WORDs (see Table 1) is 99 msec. greater than that for correct responses and similar to the overall mean for correct responses on CCVCCs. In this case it appears that a match is not found and the subset of representations is searched exhaustively. The average latency for errors on the CCVCCs is considerably less than that for correct responses and fairly close to the overall mean for correct responses on WORDs. Presumably, *S* has (mistakenly) found a match before completing his search of the subset.

Within the category of WORDs the relationship between consonant pair frequency and latency is very different than that within CCVCCs or CCCCCs. A plausible explanation is that there is another kind of frequency associated with WORDs, namely, unit frequency, and unit frequency has a quite different effect on latency than does component frequency. A finding in support of this hypothesis is the correlation between mean latency (taken over *S*s) for each WORD and its Thorndike-Lorge unit frequency of $-.65$ ($p < .001$). Another finding based on a set of filler WORDs also indicates that unit frequency affects search time. Among the filler WORDs were 13 pairs which had identical initial and terminal consonant pairs but which were different in Thorndike-Lorge frequency, e.g., GLOSS and GLASS. The high-frequency group of filler WORDs has an average latency of 649 msec. while the mean of the low-frequency group is 765 msec., $t\ (12) = 4.20$, $p < .005$. The implication of the relationship between unit frequency and latency is that search through the memory subset is ordered, the representations toward the "front" of the subset being those with high frequency of usage.

The overall results suggest an outline of a memory indexing model for the present task and materials. First, the initial and terminal consonant pairs are in some way encoded, possibly phonologically, although the results are uninformative on this point. On the basis of information from these pairs a subset of memory representations is designated or marked, and the size of the subset is positively related to the number of different words in which the consonant pairs appear. Next would follow an attempt to encode the item as a unit, phonologically, and unlawful items could be rejected at this point. If the item were phonologically lawful, a search or comparison process would ensue and continue until a match was found or until all the representations in the subset had been checked. Since search time depends on subset size, the search part of the process as well as marking would be affected by the frequency of the consonant pairs.

REFERENCES

Kirk, R. E. *Experimental design: Procedures for the behavior sciences.* Belmont, Calif.: Wadsworth, 1968.

Mayzner, M. S., & Tresselt, M. E. Tables of trigram frequency counts for various word-length and letter position combinations. *Psychonomic Monograph Supplements,* 1965, **1**, 33–78.

Rubenstein, H., Lewis, S. S., & Rubenstein, M. A. Evidence for phonemic recoding in visual word recognition. *Journal of Verbal Learning and Verbal Behavior,* 1971, **10**, 645–657.

Stanners, R. F. Language frequency correlates of rated pronunciability. *Journal of Verbal Learning and Verbal Behavior,* 1970, **9**, 373–378.

Stanners, R. F., Forbach, G. B., & Headley, D. B. Decision and search processes in word–nonword classification. *Journal of Experimental Psychology,* 1971, **90**, 45–50.

Thorndike, E. L. *Thorndike century senior dictionary.* New York: Scott, Foresman, 1941.

Venezky, R. L. A computer program for deriving spelling-to-sound correlations. Unpublished master's thesis, Cornell University, 1962.

(Received May 22, 1972)

Journal of Experimental Psychology
1973, Vol. 98, No. 1, 36–43

VISUAL PATTERN MATCHING:

AN INVESTIGATION OF SOME EFFECTS OF DECISION TASK, AUDITORY CODABILITY, AND SPATIAL CORRESPONDENCE[1]

R. S. NICKERSON[2]

Bolt Beranek and Newman Inc.,
Cambridge, Massachusetts

R. W. PEW

University of Michigan

The *S*s made three types of decisions about successively presented pairs of visual stimuli: (*a*) Do the two pairs have at least one item in common? (*b*) Do they have both items in common? (*c*) Are they identical? Both response times (RTs) and error rates indicated that the first question was the most difficult to deal with and the last the easiest. Two types of stimuli were used: letters and random dot patterns. Both RTs and error rates were much larger for the latter. Performance also varied with the way in which the items of the second (probe) pair of stimuli corresponded to those of the first (target) pair; in particular, it was sensitive to whether like items in the two displays occurred in the same or different positions. The implications of these findings for questions of serial vs. parallel processing and of visual vs. auditory memory codes are discussed.

The generic task with which we are concerned is the following. The *S* is shown a set of one or more visual stimuli that he must commit to memory. This set is erased from the visual display, whereupon a second set is presented. The *S* must make a decision based on a comparison between the item(s) of the first set and that (those) of the second. For example, he might be asked to decide whether any of the items of the first set are contained in the second set, or whether the two sets are identical (same items in corresponding positions), or whether the two sets contain the same items (but not necessarily in the same positions).

Two recurring questions relating to such tasks concern (*a*) how visually presented information is represented in short-term memory and (*b*) how that information is retrieved from memory when it must be applied to the problem of classifying other visual inputs. The former question has been raised in particular with respect to visual patterns which have highly overlearned auditory associates or "names" (e.g., letters and numerals). The latter question has frequently been couched in serial vs. parallel processing terms. When an *S* must refer to several items in memory in order to make a decision with respect to some item on a visual display, are the memory representations "scanned" one at a time in sequence or many at once in parallel? Numerous experiments bearing on these issues have been conducted recently and several reviews of the results are available (Nickerson, 1972; Smith, 1968; Sternberg, 1969).

The models that have been generated by these studies typically have been developed to accommodate results obtained with a particular experimental procedure. Sternberg (1966, 1969), for example, has proposed a model to account for data obtained when *S*'s task is to decide whether a character on a display is contained in a specified set of one or more characters in memory. According to this model, the set of memory characters is scanned serially, each memory representation being compared with the test item in turn. Bamber (1969) has proposed a model to account for results obtained when the task is to decide whether two sets of characters, one in memory and one on a display, are identical. According to this model, one stage of the process involves a comparison of the entire set of display items against the entire set of memory items simultaneously. If both of these models are correct, we are left with the conclusion that whether one scans a set of memory items serially or in parallel is at least partially dependent on the nature of the decision that he must make.

---

[1] This work was supported by the Air Force Office of Scientific Research under Contract F44620-69-C-0115. The experiment was performed while the second author was a visiting scientist at Bolt Beranek and Newman Inc. The authors wish to thank Barbara Noel and Daniel Kalikow for their assistance in collecting and processing the data.

[2] Requests for reprints should be sent to R. S. Nickerson, Bolt Beranek and Newman Inc., 50 Moulton Street, Cambridge, Massachusetts 02138.

In the interest of parsimony, it would be helpful to define a model in which one of these decisions was the outcome of one stage of a process designed to arrive at the other: The determination that two displays have one item in common could be the first step in determining whether they are identical, or, conversely, a check for identity could be the first step in determining whether they have at least one item in common. One purpose of the present experiment was to provide some data which would bear on this issue. Response times (RTs) were obtained for different decisions involving the same set of stimuli. The question of interest was whether the results would be consistent with the notion that the same basic processes underlie the performance of each of the decision tasks studied.

A second purpose of the experiment was to obtain additional data on the question of the role of auditory codability in the storage and retrieval of visually presented patterns. To this end, two sets of stimuli were used: One was composed of visual patterns with highly overlearned auditory associates (letters), the other of random dot patterns.

A problem that is encountered when trying to find simple visual stimuli that do not have names is the fact that they may *acquire* names, if they are repeatedly presented to *S*, during the course of the experiment. For this reason, the random dot patterns that comprised the second set of stimuli were generated independently for every trial. Given the generation algorithm that was used, approximately 800 million different patterns were possible; so it would infrequently happen that *S* would see the same pattern on more than one trial.

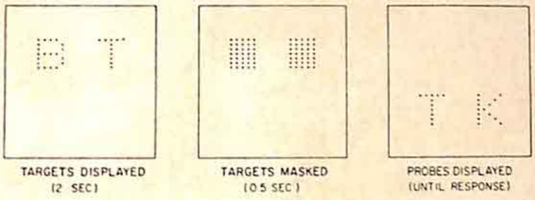


FIG. 1. An illustrative target–mask–probe sequence.

METHOD

General Procedure

Two pairs of dot patterns were displayed on a computer-controlled oscilloscope on each trial. Each pattern was formed from a subset of dots in a 7 × 5 (.95 cm. × .75 cm.) matrix. The two patterns comprising a pair always differed from each other. The first pair of patterns (target set) was masked by superimposition of 7 × 5 rectangular dot patterns before the second pair (probe set) was displayed. The two targets appeared on the display side by side, separated by a 1.11-cm. space. The two probes appeared directly under the positions of the targets and were separated from the latter by .79 cm. Targets were displayed for 2 sec., masks for .5 sec., and probes until a response was made. Figure 1 shows an illustrative target–mask–probe sequence drawn to scale. The *S*s viewed the display from a distance of approximately .6 m.

The *S*'s task was to decide as quickly as possible whether the target and probe sets were related in one of several specified ways, as explained below. Decisions were registered by pressing one of two telegraph keys. Half of the *S*s in each group pressed the left key to indicate a positive decision and half the right.

Each *S* had 28 runs (4 on the first day and 8 on each of the 3 succeeding days) under one of six conditions. The first day's runs were considered practice and data from these runs are not reported. The *S*s ran in pairs, alternating every 2 runs. Sixty trials (30 negatives and 30 positives) comprised each run.

Independent Variables

There were four independent variables of primary interest: (*a*) the type of stimulus used, (*b*) the nature of *S*'s decision task, (*c*) the relationship between target and probe sets, and (*d*) the type of response (positive or negative) required. The first two were between-*S*, variables and the latter two within-*S* variables.

Stimulus type. Two types of stimulus patterns were used: block capital letters and random patterns. It was assumed that the letter stimuli would be more readily encoded auditorily than the random patterns.

Each random pattern was composed from 12 dots positioned at random within the 7 × 5 matrix with the following constraint: Any two of the four patterns that occurred on a given trial were either identical or different with respect to at least 5 dots. The intent of this constraint, which was known to *S*, was to ensure that whatever differences occurred would be well above the discrimination threshold. Error rates were substantial even with this precaution. Scale drawings of examples of the random patterns are shown in Figure 2.

Decision task. Three different decision tasks were used: *S* had to decide in Task I whether the target and probe sets had any items in common, in Task II whether the target and probe sets had *both* items in common, and in Task III whether the two sets were identical, i.e., same patterns in the same position.

Relationships between target and probe sets. Taking into account the positions of items within their respective sets, there are seven ways in which the target and probe sets could be related. These relationships are shown in Table 1. The top two letters of any set of four represent the targets and the bottom two the probes. Note that a positive response is called for by Mapping Rules 1–6 under Decision Task I, by Rules 1 and 2 under Task II, and by Rule 1 under Task III.

Response type. For each task, half of the trials of each run called for a positive response and half for a negative. Within the constraints implied by this fact, each of the mapping rules was represented as nearly equally as possible.

Subjects

The *S*s were high-school students paid for their participation. Four different *S*s were assigned to each of the six combinations of decision task and stimulus type.

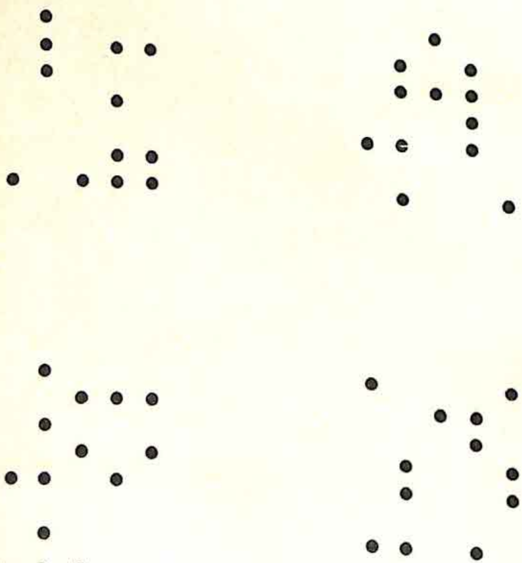


FIG. 2. Some examples of the random patterns used.

TABLE 1

SEVEN RELATIONSHIPS OF INTEREST BETWEEN TARGET AND PROBE SETS

Mapping rule no.	Target (top) and probe (bottom) sets	Symbol	Description of relationship
1	A B A B	□	Same patterns, same positions
2	A B B A	⊠	Same patterns, different positions
3	A B A C	∻	Leftmost patterns same
4	A B C B	∻	Rightmost patterns same
5	A B C A	∻	Right probe same as left target
6	A B B C	∻	Left probe same as right target
7	A B C D	○	Nothing same

RESULTS

The results were analyzed in two phases. The first phase focused on the effects of three of the experimental variables: stimulus type, decision task, and response type, along with the effect of practice. In the second phase the effects of the relationship between target and probe sets were analyzed for Day 3 data only.

Effects of Stimulus Type, Decision Task, and Response Type

A 2 × 3 × 2 × 3 analysis of variance was performed with four scores (individual *S* medians for correct RTs) in each cell. The means of the medians are shown in Table 2. The main effects of these variables are summarized in Table 3.

The effect of stimulus type, F (1, 18) = 47.19, $p < .001$, was large: The random patterns required about 60% more time to process than the letters and elicited nearly

TABLE 2

Summary Data Table

Stimulus type	Task	Correct response	Day 1	Day 2	Day 3
Letters	I	Positive	575 (3)	523 (4)	510 (5)
		Negative	723 (2)	663 (4)	619 (4)
	II	Positive	535 (5)	465 (5)	450 (5)
		Negative	542 (2)	479 (3)	479 (5)
	III	Positive	396 (4)	361 (3)	353 (3)
		Negative	431 (4)	402 (3)	398 (6)
Random figures	I	Positive	1,153 (30)	878 (27)	907 (34)
		Negative	1,398 (23)	1,062 (22)	1,003 (20)
	II	Positive	915 (16)	822 (15)	740 (17)
		Negative	1,033 (18)	888 (23)	784 (22)
	III	Positive	585 (10)	525 (7)	474 (6)
		Negative	600 (15)	547 (18)	491 (19)

Note. Each cell entry is a mean of four *S* medians (in msec.). Numbers in parentheses are percentage error rates.

five times as many errors. The effect of decision task, $F\ (2, 18) = 20.12, p < .001$, was also large, both RT and error rate indicating that, from least to most difficult, the task were ordered III, II, and I. On the average, negative RTs were about 72 msec. longer than positives, $F\ (1, 18) = 28.79, p < .001$. False positives and false negatives occurred equally frequently. Response time decreased with practice, $F\ (2, 36) = 54.97, p < .001$; however, the difference between the second and third days was relatively small. Error rates, pooled over the other variables, did not change with practice.

The following interactions were also significant: Stimulus Type × Task, $F\ (2, 18) = 3.7, p < .05$; Stimulus Type × Practice, $F\ (2, 36) = 16.60, p < .001$; Task × Response Type, $F\ (2, 18) = 7.08, p < .01$; Task × Practice, $F\ (4, 36) = 5.68, p < .005$; and Stimulus Type × Task × Practice, $F\ (4, 36) = 3.62, p < .025$. Each of the two-way interactions may be described by saying that when the value of one of the interacting variables was changed so as to decrease the difficulty of the judgment (decrease the RT), the effect of the other variable was attenuated. Similarly, the three-way interaction may be summarized by saying that the magnitude of the interaction between any two of the variables depended on the difficulty of the task as determined by the remaining variable.

Effects of the Relationship between Target and Probe Sets

Table 4 shows, for letters and random patterns separately and for each decision task, the dependence of RT and error rate on the various target-set-to-probe-set mappings that could occur.

Because practice interacted both with stimulus type and with decision task, it seemed best to confine our attention in this analysis to the data obtained on Day 3. The scores in Table 4 represent performance after one practice and two experimental sessions (20 runs).

TABLE 3

Summary of Main Effects

Variable	Level	RT[a]	Error rate (%)
Stimulus type	Letters	493	4
	Random figures	812	19
Task	I	818	15
	II	677	11
	III	461	8
Correct response	Positive	616	11
	Negative	688	12
Day	1	740	11
	2	635	11
	3	600	12

[a] RT is the mean of medians (in msec.).

TABLE 4

EFFECTS OF THE RELATIONSHIP BETWEEN TARGET AND PROBE SETS ON RT AND ERROR RATES

Mapping rule no.	Symbol	Decision task		
		I	II	III
		Letters		
1	□	438 (2)	407 (3)	353 (3)
2	⊠	478 (2)	499 (8)	432 (9)
3	:+	503 (4)	500 (8)	430 (14)
4	+:	528 (3)	515 (7)	398 (9)
5	˙+.	584 (11)	485 (7)	399 (2)
6	.+˙	593 (11)	507 (8)	390 (2)
7	○	619 (4)	444 (1)	373 (3)
		Random figures		
1	□	744 (10)	653 (7)	474 (6)
2	⊠	992 (22)	936 (27)	488 (16)
3	:+	840 (39)	883 (31)	547 (45)
4	+:	854 (36)	836 (40)	524 (32)
5	˙+.	970 (50)	814 (21)	496 (8)
6	.+˙	940 (44)	826 (21)	484 (8)
7	○	935 (20)	720 (8)	480 (7)

Note. Data are for Day 3 only. RTs are means of *S* medians (in msec.). Error rates, rounded to the nearest percent, are in parentheses. The horizontal lines divide each set of numbers into positive and negative RTS, the numbers above the lines representing positives.

In general, observations that are made regarding ordinal relationships between subsets of scores within this table pertain to both the top and bottom halves. The rank-order correlations between corresponding scores in the top and bottom halves of the table are .85 for RTs and .55 for error rates. The most striking exceptions to the ordinal consistency across these sets involve the RTs associated with Decision Tasks I and II and Mapping Rule 2. In these cases, the random patterns took unusually long to judge when target and probe sets had both patterns in common but in opposite positions.

Response times and error rates were also positively correlated, the rank-order correlations in this case being .37 for letters and .68 for random patterns. Thus, we may assume that the ordinal relationships between RTs that are observed are not attributable to speed–accuracy trade-offs.

DISCUSSION

Stimulus Type

It is tempting to attribute the large effect of stimulus type to the availability of auditory codes for the letter stimuli. Such a conclusion would be consistent with the generally accepted notion that visual and auditory short-term memories (STM) decay at different rates and that verbal material can be kept accessible in auditory STM through rehearsal, provided the amount of material is not too great. From this viewpoint, the random-pattern targets would decay rapidly in visual STM and, because they have no names, would not be accessible from auditory STM as would the letters. We will consider the tenability of this explanation in the light of the importance of position effects that are discussed below.

Decision Task

Apparently, it was easier to decide whether two pairs of stimuli were "identical" than to decide whether they had the same items irrespective of position, which in turn, was easier than to decide whether they had at least one item in common. To assess better the effect of decision task, we should consider the two mapping rules (1 and 7) that called for the same response (positive and negative, respectively) in all three cases. An inspection of the RTs associated with these rules (Table 4) reveals that the time it took to make a *particular response* to a *particular stimulus* depended in large measure on the nature of the decision required. When the target and probe displays had nothing in common (Rule 7), negative RT was longest when the question was, "Do they have at least one item in common?" (Task I) and shortest when it was, "Are the displays identical?" (Task III). This is what one would expect if the decisions in all cases were based on self-terminating item-by-item comparisons of the displays (whether serial or parallel), since four comparisons would be required with Decision Task I, two with Task II, and one with Task III. However, positive

RTs for the case in which the target and probe sets were identical (Rule 1) are ordered in the same way. This result is not consistent with the assumption of a self-terminating item-by-item comparison process, inasmuch as Task I would require fewer comparisons than either II or III. How many comparisons would be required by a serial process on any given trial would depend on the order in which the comparisons are made. Unless that order covaried with the nature of the stimuli, however, fewer comparisons would be required by Task I, on the average, than by II and III.

If the process were fully parallel—all possible comparisons made at once—Task I would be expected to yield the shortest RTs (given Mapping Rule 1) because the response could be triggered as soon as the first match is obtained; whereas in both the other cases, the response would have to wait until two matches had been found. Other self-terminating partially parallel processes (e.g., one probe item at a time compared simultaneously with both targets, one target compared simultaneously with both probes, or both probes compared simultaneously with the targets occupying corresponding positions) also would lead us to expect Task I to yield the shortest RTs (again, with Mapping Rule 1). The fact that the shortest RTs are actually obtained with Task III is problematic for any model that assumes that a common item-by-item comparison process underlies the performance of all three tasks.

The importance of the decision task is also indicated by the RTs obtained with Mapping Rules 3–6 under Decision Tasks II and III. Again, these tasks required the same response (a negative one) with each of these mapping rules. It appears, however, that given target and probe sets with only one item in common, the time required to decide no was greater when the question was, "Do the sets have both items in common?" than when it was, "Are they the same?" In the cases of Mapping Rules 5 and 6, this result might be attributed to the fact that these patterns come closer to satisfying the criterion for a positive response under Task II than under Task III. This is not true, however, of Mapping Rules 3 and 4. Both of these patterns come as close to satisfying the criterion for a positive response under one of these tasks as under the other; nevertheless, RT is much faster under Task III.

The conclusion that we draw is that the strategy that *S* adopts for dealing with a specific stimulus depends strongly on the nature of the judgment that he is required to make. In particular, he may do something qualitatively different when called upon to decide whether two displays are identical than when required to determine whether those same displays have any common components. The data seem to rule out the possibility that the same process is being used consistently across tasks. They support the view that as an information processor the human being has a varied repertory of procedures that he may apply selectively to processing tasks and that his performance strategy in any particular situation will depend not only on the nature of the stimuli with which he must deal but also on the details of the decision that he must make.

Importance of Position

In this section several comparisons are made between RTs obtained with different mapping rules. Consider first Mapping Rules 1 and 2 and Decision Tasks I and II. Both of these rules called for positive responses with both tasks, and in each case the stimulus and probe sets had both items in common. What is clear from a comparison of RTs associated with the two rules is the importance of position: RT was considerably shorter when the matching target and probe items occupied the same positions in their respective displays. One-tailed median tests were performed on the data from each *S*, comparing the medians obtained with Mapping Rules 1 and 2. Of the 16 tests (4 *S*s in each of the 4 relevant combinations of stimulus type and decision task) 13 were significant (11 at $p < .01$; 2 at $p < .05$). This rules out the possibility of a totally parallel, position-insensitive process in which every probe item is compared with every target item simultaneously. Such a process would be logically inadequate for Decision Task III, inasmuch as position information is relevant to the decision that is required in that case. With Decision Tasks I and II, however, position information was irrelevant; nevertheless, it apparently was not ignored.

The fact that, under Decision Task I, RT was faster with Mapping Rule 1 than with Rule 2 also is further evidence against the two self-terminating, partially parallel processes mentioned above in the discussion of stimulus type effects, namely processes that are parallel with respect to the targets and serial with respect to the probes, or parallel with respect to the probes and serial with respect to the targets. On balance, the differences between the RTs obtained with Mapping Rules 1 and 2 under Decision Tasks I and II are suggestive of a comparison procedure that includes something akin to a test for congruence between the two stimulus sets as wholes (Bamber, 1969). The fact that the differences were much smaller for letters than for random patterns further suggests that whatever process or processes may be used in addition to congruence testing are more efficient for letters.

Given the importance of position in those cases in which target and probe sets contained the same items, we might expect to find some evidence of a similar effect in those cases in which the two sets had only a single item in common (Mapping Rules 3–6). If a match between two items is more readily detected when those items occupy the same positions in both target and probe sets, we would expect matching items in corresponding positions to have a facilitating effect when a positive response is called for and a detrimental effect when the required response is a negative one. Thus, with Decision Task I, Mapping Rules 3 and 4 should, on the average, produce shorter RTs than Rules 5 and 6. In contrast, with Decision Tasks II and III, Rules 3 and 4 should produce the longer RTs. Inspection of Table 4 reveals that the data are generally consistent with this expectation.

Error patterns give further evidence of the importance of position. First, in the cases of Decision Tasks I and II, Mapping Rule 2 generally produced a greater incidence of false negatives than did Rule 1. Second, consider the error patterns obtained in those cases in which target and probe sets had a single item in common. With Decision Task I, in which a single match was sufficient to justify a positive response, the probability of a false negative was greater when the common item did not occupy the same position in both sets than when it did. With Decision Task III, in which a single matching item called for a negative response, the probability of a false positive was higher when the matching item occupied the same position in both sets than when it did not. The error pattern obtained with Decision Task II shows no position effect for letter stimuli, but is similar to that obtained with Task III and thus consistent with the assumed effect for random patterns.

Position effects such as these are more readily accounted for if the stimulus representations are assumed to be visual in nature than if they are thought of as exclusively auditory codes. This is not to suggest that position effects could not be obtained with auditorily encoded stimuli; any serial self-terminating process that scanned either a physical or a memorized display in a consistent order might produce them. It is difficult to see, however, any compelling reason for it to be particularly advantageous (or disadvantageous) for matching target and probe items to occupy the *same* position within the display, if the stimulus representations were exclusively auditory. We conclude that the letter stimuli, as well as the random patterns, used in this experiment were represented in memory in a "visual-like" form and that the decisions that were required were based, at least in part, on comparisons involving those visual-like representations.

Letters vs. Random Figures

We see three possible explanations of why performance was faster and more accurate with the letters than with the random figures: (*a*) the greater familiarity of the letters, (*b*) the greater complexity or lesser discriminability of the random patterns, and (*c*) the dual encoding of the letters in memory. All three possibilities are plausible, as is the possibility of some combination of these factors.

Consider first the possible role of stimulus familiarity. Posner and Mitchell (1967) have presented data suggesting that familiarity with the stimuli is not a significant determinant of the time required to match *simultaneously presented*, letter-like forms. As the comparison that is required becomes more memory dependent, however, familiarity does play a role (Hochberg, 1968; Posner, 1969). In the present experiment the temporal separation of the items to be compared insured the implication of memory, so a familiarity effect might be expected. However, if the memory representation of the letter stimuli was at least partially visual, as we have concluded, the question arises concerning how familiarity might help to keep such a representation intact. If *S* can rehearse or generate (Posner, Boies, Eichelman & Taylor, 1969) a visual image, it seems reasonable to assume that he should be able to do so more effectively with patterns that are familiar than with those that are not. An unfamiliar pattern once forgotten may be irretrievably gone; a familiar one, particularly if its name is remembered, may be reconstructed even after its trace has faded. Or, perhaps the trace of a familiar pattern is more maintainable simply because one knows what it is that must be maintained.

It is undoubtedly true that the letters *were* more familiar to *Ss* in this experiment than were the random patterns. One wonders to what extent familiarity and auditory codability are in principle separable when the stimuli are simple visual patterns. A pattern that is familiar is almost bound to have a name; and, conversely, one that has a readily accessible name is by definition familiar.

The second possible explanation of the superior performance with the letters is that the two stimulus types differed in complexity or in "goodness," in the sense in which Garner (1966) has used the word. The patterns resulting from the random positioning of dots in a 7 × 5 matrix are probably less likely than the letters to be comprised of continuous curves and connected line segments or to display the symmetries and closures that would characterize "good" forms. The relative importance of complexity or goodness might be better assessed by using as random figures random combinations of letter components joined in such a way as to configure letter-like shapes or to show the organizational properties associated with good shapes.

The possibility that stimulus discriminability was an important factor in determining these results might be tested by an experiment in which the interitem discriminability of the random dot patterns is varied systematically. One would like to know whether more easily discriminated patterns would yield results more like those obtained with letters.

Finally, there is the possibility that the letters (but not the random patterns) were represented in memory in *both* visual and auditory form. If one assumes that a name-comparison process occurs concurrently with a process that compares visual representations, and that a response can be initiated by the completion of either process, then one has another basis for accounting for the relative speed of performance with the letter stimuli. This set of assumptions also can account for the fact that the difference between RTs for the letters and the random patterns is least when the question that must be answered involves physical identity of the two displays (Decision Task III) and especially great when a positive response is called for by stimuli that would fail most completely on a holistic congruence test (Mapping Rule 2 under Decision Tasks I and II and Rules 5 and 6 under Task I).

Conclusions

The results of this experiment are not fully accounted for by any existing model of character classification. They support the following tentative conclusions, however, about the nature of the processes involved in the performance of these tasks: (*a*) The strategy that was used to compare two sets of stimuli differed depending on the nature of the decision task; (*b*) both the letters and the random patterns were represented in memory in a visual-like form; (*c*) the letters were probably also represented auditorily; (*d*) under some conditions the decisions were based on holistic congruence matching of the visual patterns; and (*e*) the faster processing of letters was a result of a combination of their greater familiarity, their lesser complexity or greater discriminability, and their dual representation in memory.

REFERENCES

BAMBER, D. Reaction times and error rates for "same"-"different" judgments of multidimensional stimuli. *Perception & Psychophyscis*, 1969, **6**, 169-174.

GARNER, W. R. To perceive is to know. *American Psychologist*, 1966, **41**, 11-19.

HOCHBERG, J. In the mind's eye. In R. N. Haber (Ed.), *Contemporary theory and research in visual perception*. New York: Holt, Rinehart & Winston, 1968.

NICKERSON, R. S. Binary-classification reaction time: A review of some studies of human information-processing capabilities. *Psychonomic Monograph Supplements*, 1972, **4**, (17, Whole No. 65).

POSNER, M. I. Abstraction and the process of recognition. In K. W. Spence & G. Bower (Eds.), *Psychology of learning and motivation*. Vol. III. New York: Academic Press, 1969.

POSNER, M. I., BOIES, S. J., EICHELMAN, W. H., & TAYLOR, R. L. Retention of visual and name codes of single letters. *Journal of Experimental Psychology Monograph*, 1969, **79** (1, Pt. 2).

POSNER, M. I., & MITCHELL, R. F. Chronometric analysis of classification. *Psychological Review*, 1967, **74**, 392-409.

SMITH, E. E. Choice reaction time: An analysis of the major theoretical positions. *Psychological Bulletin*, 1968, **70**, 77-110.

STERNBERG, S. High-speed scanning in human memory. *Science*, 1966, **153**, 652-654.

STERNBERG, S. Memory-scanning: Mental processes revealed by reaction-time experiments. *American Scientist*, 1969, **57**, 421-457.

(Received May 25, 1972)

Journal of Experimental Psychology
1973, Vol. 98, No. 1, 44–48

FUNCTIONAL INDEPENDENCE OF PICTURES AND THEIR VERBAL MEMORY CODES[1]

DOUGLAS LEE NELSON[2] AND DAVID HEBER BROOKS

University of South Florida

Phonetic similarities among paired-associate (PA) stimuli interfere with rate of acquisition independently of the imagery value of the stimulus words. The purpose of this experiment was to determine if phonetic attributes play a role in coding when the information processing sequence is initiated from a pictorial rather than a verbal representation. The Word condition contained stimuli which were either high or low in similarity, and the Picture condition consisted of pictorial representations of these words. Half of the *S*s in the Picture condition were given standard PA instructions and the remaining half were, in addition, required to overtly name each picture before anticipating its response. Similarity impaired acquisition to the greatest extent within the Word condition, to a lesser extent in the Picture–Name condition, and not at all in the Picture condition. In conjunction with verbal report data, these findings suggest that pictorial representations can function as effective memory codes independently of their corresponding verbal representations.

In a recent series of experiments, Nelson and Brooks (1973) have shown that the phonetic features of words apparently have as great a functional value for processing high-imagery concrete words as for processing low-imagery abstract words. Phonetic similarity and imagery level were manipulated within the stimulus words of paired-associate (PA) lists. The dependence of word coding processes upon phonetic features was implicated by the interference produced by the sharing of common elements. The dependence of these processes upon imaginal features was implicated by differential acquisition rates associated with concrete and abstract words and by the effects of instructional manipulations. Most importantly, however, the sharing of phonetic features impaired performance independently of imagery level: Phonetic similarity among concrete words produced increments in difficulty comparable to those obtained with abstract words (Nelson & Brooks, 1973, Experiments I–III). Furthermore, these impairments were obtained even when the PA pairs were high-imagery words which were presented at a relatively slow rate following extensive practice at generating interacting images (Experiment V). Apparently, even under relatively ideal conditions for processing nonverbal features, phonetic attributes were processed as well.

The results of these experiments suggested that both the verbal (phonetic) and the nonverbal (imaginal) features of the high-imagery stimulus words were utilized as cues for processing information. The purpose of the present experiment was to determine if the coding of phonetic attributes was dependent upon initiating the information processing sequence from a verbal stimulus. Thus the major variable in this study was type of representation, verbal or pictorial. In the Word condition, the PA stimuli were phonetically similar words in one list and phonetically dissimilar words in another. The expectation for these conditions was that similarity would impair rate of acquisition even though the items were all high-imagery words. This result would of course be consistent with previous findings (Nelson & Brooks, 1973). In the Picture condition, the stimuli consisted of pictorial representations of the corresponding words. Therefore, although the verbal labels for the pictures varied in similarity, the pictorial representations were as distinct in the high-similarity condition as they were in the low-similarity condition. Given the usual superiority of pictures over words as PA stimuli (e.g., Paivio, 1971; Wicker, 1970), it was expected that rate of acquisition would be faster in the Picture than in the Word condition when low-similarity stimuli were used. When high-similarity stimuli were used, three alternative outcomes seemed plausible, all depending upon whether or not the pictures were implicitly labeled. First, if the pictures were covertly named as they were presented, one possibility was that difficulty of acquisition would be as great for the pictorial representations as for the words when the labels were similar. This outcome would suggest that verbal labels elicited by the pictorial stimuli played the major role in directing storage and retrieval processes. Second, given implicit naming, the interference generated by similarity may be less, relative to that obtained when the labels themselves serve as stimuli. In this case, an interaction between type of representation and similarity would be expected, and this result would be consistent with the hypothesis that, given pictorial stimuli, both nonverbal and verbal features were being processed. The third alternative assumes that the pictures are not labeled. If so, then pictorial representations should be associated with faster acquisition than words and label similarity should not interfere with acquisition. This result, which would also be reflected in a Representation × Similarity interaction, would suggest that nonverbal representations can function effectively as memory codes independently of their verbal counterparts. Given the plausibility of the latter outcome, a third set of conditions was included in which *S*s were required to overtly name each picture as it was shown. This requirement was expected to generate a phonetic interference effect because of the similarity among the labels in the high-similarity condition. However, to the extent that imaginal features are also processed in this condition it was expected that the interference effect might be less than that obtained in the Word conditions.

---

[1] This research was supported by Grant MH-16360 to the first author from the National Institute of Mental Health, U. S. Public Health Service.

[2] Requests for reprints should be sent to Douglas Lee Nelson, Department of Psychology, University of South Florida, Tampa, Florida 33620.

METHOD

Materials. Each *S* acquired a single list of six PAs in which the stimuli were either words or their pictorial representations. In different lists, the verbal stimuli were all relatively common six letter words which were either high or low in formal similarity. The pairs for the two high-similarity lists were: BASKET–MARK, BONNET–EDGE, BULLET–FORK, BEAVER–DOOR, BUTLER–HARP, BUMPER–VEST for List 1; and PARROT–MARK, PEANUT–EDGE, PUPPET–DOOR, PADDLE–FORK, PUZZLE–VEST, POODLE–HARP for List 2. Thus high-similarity stimuli shared the same first and last letters. The pairs for the two low-similarity lists were: BASKET–MARK, PADDLE–FORK, CACTUS–EDGE, GARDEN–VEST, TURKEY–DOOR, SPIDER–HARP for List 1; and BULLET–FORK, PUZZLE–VEST, CANNON–DOOR, WINDOW–MARK, ANCHOR–HARP, MONKEY–EDGE for List 2. Low-similarity items shared a minimum of letters. Response terms for all lists were high-imagery relatively common four-letter words taken from the Paivio, Yuille, and Madigan (1968) norms. All responses were randomly assigned to stimuli with the restriction that no obvious associations were formed. The pictorial stimuli consisted of line drawings representing, respectively, the high-similarity and the low-similarity stimuli. Thus, the verbal labels for the pictures varied in similarity but the pictorial representations were apparently equally distinctive in each similarity condition.

Procedure. Standard PA instructions were given to all *S*s. Those in the Picture and those in the Word conditions were instructed to vocally anticipate the response before the pair appeared on the screen. The *S*s in the Picture–Name condition were instructed to overtly articulate the name of the picture stimulus before anticipating its response, and to aid them in doing so, *E* named each picture as it was shown on the study trial. A comparable overt-articulation condition for word stimuli was not included since previous findings have suggested that manipulation of this variable is ineffective under these task conditions (Nelson, Brooks, & Fosselman, 1972).

In individual sessions, practice was by the method of anticipation for a total of 20 trials following the study trial or for 4 consecutive errorless trials, whichever came first. If the latter criterion was achieved, the protocol was scored as if there were no additional errors. Slides of all items (black on light background) were presented via a Kodak Carousel slide projector driven at a 1.5:1.5 sec. rate. There was no intertrial interval. Sequences of pairs were presented in five unsystematically varying orders, with the restriction that at least two pairs had to intervene between the presentation of a pair and its recurrence. This sequencing was changed for each *S* in each condition.

All *S*s learned a practice list consisting of two pairs comparable to those used in the experimental task. Immediately following the experimental task each pair was shown individually. In the Picture condition, *S* was first asked to name the picture, then to describe how the pair was learned, and finally, if he

TABLE 1

MEAN TOTAL ERRORS AS A FUNCTION OF TYPE OF REPRESENTATION AND SIMILARITY

Similarity	Type of representation		
	Picture	Picture–Name	Word
High	17.81	40.19	69.75
Low	21.44	33.31	39.56

thought of the name of the picture while learning the list. In the Picture–Name and Word conditions, each *S* was asked to describe how each pair was learned. The *E* classified each description into one of three categories: rote, verbal mediation, or imagery mediation.

Subjects. With two conditions of similarity, three types of representation (Picture, Picture–Name, and Word), and two lists, there were a total of 12 different between-*S* conditions. Eight *S*s were assigned to each of these treatment combinations giving a total of 96 *S*s. These *S*s were assigned to conditions in blocks of 12, with 1 *S* from each condition per block. Assignment within blocks was determined by a table of random numbers. All *S*s were selected from courses in introductory psychology and received points toward their grade for participation. A single male *E* collected all data.

RESULTS

A trial-by-trial analysis of variance of errors to the acquisition criterion (omission plus commission) indicated that the main effects of similarity, F (1, 90) = 9.08; type of representation, F (2, 90) = 29.91; and rials, F (19, 1710) = 177.19, were all tatistically significant at $p < .01$. Mean total errors per *S* for low and high similarity were, respectively, 31.44 and 42.58, and the mean values for the Picture, Picture–Name, and Word conditions were 19.62, 36.75, and 54.66, respectively. Thus similarity impaired performance, and the Picture–Name condition was intermediate in difficulty relative to the Picture and Word conditions. This analysis also revealed a reliable Similarity × Type of Representation interaction, F (2, 90) = 7.30, $p < .01$, indicating that label similarity had differential effects within types of representation. Mean total errors for this interaction are shown in Table 1. Fisher's two-tailed least significant difference for subsequent test comparisons on this interaction was 2.84. Thus, all means were reliably different from each other. Comparison of these values indicated that the differences between high and low similarity were greatest within the Word condition and next greatest within the Picture–Name condition. The direction of this difference was actually reversed in the Picture condition. Thus, similarity impaired performance to the greatest extent when storage and retrieval processes were initiated from a word, next when the initiation was from a picture which had to be labeled, and apparently not at all when *S*s were not required to overtly label the pictorial representation. It should also be noted that even when label similarity was low, preformance in the Picture–Name condition was inferior to that in the Picture condition, suggesting that the act of overtly articulating the label for the picture interfered with performance. The only other reliable interactions in this analysis were Type of Representation × Trials, F (38, 1710) = 3.63, $p < .01$, and Similarity × Type of Representation × Trials, F (38, 1710) = 1.59, $p < .05$. The latter interaction reflected the persistance of interference throughout the acquisition period when the stimulus items were similar words.

Upon reaching the acquisition criterion, *S*s in the Picture condition were required to label each of the stimuli. It was assumed that these responses could be used to determine the degree of concordance between *E*s and *S*s as to the most appropriate label for each picture. This agreement was generally high with correct labeling obtained for an average of 4.84 out of the 6 stimuli. Disagreement appeared consistently on two items: butler–"a man holding a coat" and paddle–"an oar." When asked if the name for the picture occurred to them during the learning phase, *S*s reported that labels were used for an average of 1.25 items. However, many qualified this statement by indicating that it occurred to them only at the very beginning of the task. The majority of *S*s reported that they simply knew what the picture was and that it did not occur to them to name it. Thus, of the 32 *S*s in this condition, 12 reported never naming any picture, 10 reported naming a single picture, 5 reported naming 2, and 5 reported naming 3–6 of them. Furthermore, reported labeling did not vary as a function of label similarity, with an average of 1.37 for high similarity and 1.12 for low similarity ($F < 1$).

Separate analyses of variance were performed within each of the three strategy classifications. For both the rote and verbal-mediation classifications neither similarity, type of representation, nor the interaction of these sources even approached statistical reliability. The mean number of pairs learned, out of six classed as being rote, was 3.61, and the mean number of pairs upon which verbal mediation was reportedly used was 1.11. There did appear to be a tendency for the utilization of imagery mediation to vary with type to representation: For the Picture, Picture–Name, and Word conditions the mean number of pairs reportedly involving this strategy was 1.53, 1.22, and .78, respectively. However, an analysis of variance comparable to those used on the other classifications revealed that this source did not attain statistical significance, $F(2, 90) = 2.96$, $p < .10$.

DISCUSSION

The acquisition results of this experiment have replicated findings typically obtained when type of representation and similarity have been varied among PA stimuli. Pictures were associated with faster acquisition relative to their corresponding verbal representations, and similarity among the verbal stimuli interfered with performance. Most importantly, however, label similarity within the pictorial representations did not interfere with acquisition unless the task explicitly required an overt naming response. The null effect of similarity within the Picture condition suggests that easily recognized nonverbal representations of meaningful phenomena can function effectively as memory codes independently of their corresponding verbal representations. The verbal report data appear to corroborate this inference. Although verbal labels for most of the pictures were readily supplied upon request, the great majority of *S*s reportedly did not name any of the stimuli or utilized a label for only one of them during acquisition. Furthermore, the labels that were provided suggested that *S*s were not implicitly labeling components of the pictures, but that they were being coded wholistically, as "gestalts." In fact, nothing in the report data collected during the remainder of the interview suggested that *S*s were associating responses to specific components of the pictorial representations. Thus, the null effect of similarity cannot be attributed simply to implicit labeling of isolated features with a consequent reduction or elimination of the similarity variable. This inference is not intended as a claim that such component processing only rarely or never occurs, for there is ample evidence that it apparently does, at least when the nonverbal stimuli are complex nonsense forms (e.g., Clark, 1965; Ellis, Muller, & Tosti, 1966). The only claim that is necessary is that under the present conditions, with static and highly familiar pictorial representations, such feature sampling apparently did not occur.

The findings of this experiment in conjunction with those reported by Nelson and Brooks (1973) suggest that the processing of the phonetic attributes of PA stimuli are highly dependent upon the type of representation from which the information processing sequence is initiated. Thus, if this initiation is from a verbal representation, the evidence suggests that both the phonetic and the imaginal features of the stimulus words are processed. However, if it is from a pictorial representation, involvement of the verbal system apparently is an unnecessary requirement, unless the task specifically demands participation of this system. In the latter case, as with words alone, both features appear to be processed. Thus performance in the Picture–Name condition was intermediate in difficulty relative to the Picture and Word conditions. These results are consistent with the inference that imaginal and phonetic features are apparently processed independently, perhaps involving separate systems (e.g., Bower, 1970; Brooks, 1968; Paivio, 1971). The relative superiority of pictures over words obtained in this experiment might best be attributed to associational factors such as those postulated by the conceptual–peg hypothesis (Paivio, 1971) or to stimulus-differentiating factors such as those presumed by the discrimination hypothesis (e.g., Wicker, 1970). Given the apparent minimal utilization of interaction imagery obtained in this study, the present results favor the latter alternative, but clearly do not eliminate the former.

In addition to suggesting that the pictures were not implicitly labeled, the results of this experiment indicate that requiring overt identification prior to response anticipation interfered with performance in this task. This effect is most clearly demonstrated by the increment in difficulty observed in the Picture–Name condition relative to the Picture condition, when similarity was low. Bahrick and Boucher (1968) have observed a similar interference effect attributable to overt labeling. Relative to a group left to their own devices, they found that naming pictures of familiar objects over repeated trials had a detrimental effect on both an immediate and a delayed recognition test. In the present study, several innocently recalcitrant *S*s occasionally anticipated the response at the appearance of the test picture and then labeled the picture, rather than the reverse, but prescribed, procedure. This phenomenon appeared relatively infrequently, with not more than seven or eight *S*s exhibiting it on not more than eight occasions throughout the learning phase of the experiment. Furthermore, it appeared equally within conditions of low and high label similarity. However, this phenomenon has never been previously observed within a single *S* among the thousands of *S*s who have participated in experiments in this laboratory under conditions requiring overt articulation of word stimuli, and hence, it attests to the apparent reluctance of *S*s to perform what they reportedly regard as an unnecessary and interfering task. To account for the increase in difficulty produced by the naming requirement in the low-similarity condition, it might be argued that the labels ascribed by *E* to the pictures on the study trial were incongruent with those preferred by *S*s. However, this alternative seems unlikely since the labels reported in the Picture condition were high in apparent agreement with those of *E*s. A more plausible explanation is that forced involvement of the verbal system may have reduced the time available for coding the superior pictorial representation. Accordingly, the differential difficulty observed in the high label similarity Picture–Name condition apparently reflects the combined detrimental effects of the overt articulation requirement and the interference generated by the similarity of the verbal representations.

REFERENCES

Bahrick, H. P., & Boucher, B. Retention of visual and verbal codes of the same stimuli. *Journal of Experimental Psychology*, 1968, **78**, 417–422.

Bower, G. H. Analysis of a mnemonic device. *American Scientist*, 1970, **58**, 496–510.

Brooks, L. R. Spatial and verbal components of the act of recall. *Canadian Journal of Psychology*, 1968, **22**, 349–368.

Clark, H. J. Recognition memory for random shapes as a function of complexity, association value, and delay. *Journal of Experimental Psychology*, 1965, **69**, 590–595.

Ellis, H. C., Muller, D. G., & Tosti, D. T. Stimulus meaning and complexity as factors in the transfer of stimulus predifferentiation. *Journal of Experimental Psychology*, 1966, **71**, 629–633.

Nelson, D. L., & Brooks, D. H. Independence of phonetic and imaginal features. *Journal of Experimental Psychology*, 1973, **97**, 1–7.

Nelson, D. L., Brooks, D. H., & Fosselman, J. R. Words as sets of features: Processing phonological cues. *Journal of Experimental Psychology*, 1972, **92**, 305–312.

Paivio, A. Imagery and verbal processes. New York: Holt, Rinehart, and Winston, 1971.

Paivio, A., Yuille, J. C., & Madigan, S. Concreteness, imagery and meaningfulness values for 925 nouns. *Journal of Experimental Psychology Monograph Supplement*, 1968, **76**(1, Pt. 2).

Wicker, F. W. On the locus of picture-word differences in paired-associate learning. *Journal of Verbal Learning and Verbal Behavior*, 1970, **9**, 52–57.

(Received May 27, 1972)

Journal of Experimental Psychology
1973, Vol. 98, No. 1, 49–54

VISUAL INFORMATION AND REDUNDANCY IN READING[1]

MURIEL C. THOMPSON AND DOMINIC W. MASSARO[2]

University of Wisconsin

Previous studies have demonstrated that a letter is better identified when embedded in a valid spelling pattern than when presented alone. The present results replicated earlier findings in a paradigm that controls for redundancy by presenting the response alternatives after the stimulus presentation. However, this study failed to find an effect of visual similarity. This result suggests that presenting the alternatives right after the stimulus presentation does not control for redundancy since alternatives are eliminated during perception of the stimulus, not after the visual information has been completely analyzed. Identification was, therefore, examined in a task in which *S*s were given the letter and word alternatives prior to the experimental session. This task eliminated the spelling pattern advantage and also provided evidence for visual similarity effects in letter and word identification. The findings are consonant with a visual identification process that analyzes visual features of letters in both letter and word recognition.

Although the study of reading has had a long history (Cattell, 1886; Huey, 1908; Woodworth, 1938), there is no generally acceptable account of the processes involved in letter and word recognition. One possible reason for this may be the confounding of visual and higher order information in many visual recognition experiments. With efforts aimed at studying the visual information available in a visual display, investigators have not adequately controlled for higher order psychological processes. Consequently, inferences drawn from these studies are based on results confounded by an interaction of at least two processes.

Identifying the processes that are involved in discriminating and classifying a visual configuration requires a distinction between two sources of information available to the reader. The first consists of perceptual information that is directly available from the visual configuration of letters and words. The second source is the reader's knowledge of the valid letter sequences that can occur in words. This source of higher order information (sometimes called redundancy) reduces the possible number of valid alternatives for each letter of a word. Knowing the constraints which characterize letter sequences supplements the visual information, thus greatly facilitating the reading process (Bruner & O'Dowd, 1958; Miller, Bruner, & Postman, 1954).

The present research studies how visual information and redundancy contribute to letter and word identification. The first goal was to determine the size of the visual or perceptual unit employed in word identification. A perceptual unit is defined as a visual configuration that is uniquely represented by a property list in long-term memory, where the property list describes the set of visual features in the perceptual unit. The two main theories of the perceptual unit are the single-letter and letter-cluster theories, respectively.

The single-letter theory assumes that word recognition proceeds in terms of an analysis of the visual features of each letter in the word. The identity of the word follows analysis of the visual features in each of the individual letters. The letter-cluster theory is based on the assumption that spelling patterns or words are read without first analyzing the features of their individual letters. As a primary proponent of a letter-cluster viewpoint, Eleanor Gibson (Gibson, Pick, Osser, & Hammond, 1962) hypothesized that the functional unit in word recognition is a group of letters forming a spelling pattern. A spelling pattern is a letter cluster that has a specific and relatively invariant spelling-to-sound correlation. Gibson asserts that spelling patterns will be perceived more easily than random letter strings, because in terms of the framework given here, spelling patterns can function as perceptual units, whereas random letter strings cannot. In this case, identification of a spelling pattern would simply involve identification of one perceptual unit, whereas identification of random letter strings involves the identification of two or more units. It should be stressed that according to the letter-cluster theory, the spelling pattern is assumed to have a perceptual advantage in addition to the advantage afforded by the operation of redundancy.

---

[1] This research is based on a master's thesis carried out by the first author under the direction of the second author.

[2] Requests for reprints should be sent to Dominic W. Massaro, Department of Psychology, University of Wisconsin, Madison, Wisconsin 53706.

Accordingly, a critical test between these theories requires that the contribution of redundancy in the spelling pattern be eliminated. Reicher (1969) has recently provided such a test by comparing letter and word recognition while attempting to hold the effects of redundancy constant. In his study, *S*s were presented with either one or two letters, one or two four-letter words, or one or two four-letter nonwords. Immediately after the stimulus flash, a visual noise mask was presented that covered the former stimulus position. The masking card also presented two letter alternatives, one of which appeared in the original stimulus and one which was incorrect. To eliminate the operations of redundancy on word trials, both letter alternatives formed a common word with the other letters of the word. The results indicated that *S*s were more likely to choose the correct alternative when the stimulus was a word than when the stimulus was a letter or a nonword.

If redundancy or some other process cannot account for the fact that a letter is better identified in a spelling pattern, the perceptual unit must be larger than an individual letter. According to our definition of a perceptual unit, this means that the visual features of the unit cut across a number of letters. For example, whole-word shape or letter-cluster shape could be a visual feature. These features would differ significantly from the visual features which characterize individual letters. Accordingly, the letter and letter-cluster theories differ with respect to what features are extracted from the visual configuration.

It is therefore necessary to determine whether the same visual features are used in letter and word recognition. The literature contains sufficient data (Clement & Carpenter, 1970; Gibson & Yonas, 1966a, 1966b; Kaplan, Yonas, & Shurcliff, 1966; Neisser, 1964, 1967) which define the visual features employed in single-letter identification. This progress has resulted from Eleanor Gibson's (Gibson, Osser, Schiff, & Smith, 1963) identification of the distinctive features used in recognition of individual letters. Gibson et al. designed a list of distinctive features that describe each grapheme of the set of Roman capital letters. This feature list was then used to predict confusions between the letters and to devise an error matrix of confusions. A series of studies demonstrated that confusion errors correlated with the percentage of features that pairs of letters had in common. Other studies of letter identification (e.g., Clement & Carpenter, 1970) have shown that the letter can function as the perceptual unit, and distinct visual features of the letter are critical for letter recognition.

If the letter is always the perceptual unit of analysis, feature testing in word recognition should operate in the same way as in letter recognition. However, evidence to date contradicts this assumption. The findings of a word advantage effect strongly suggest that feature extraction operates on the spelling pattern of the entire word rather than on the individual letter. In the present study, the visual similarity of letters was varied to test the nature of the perceptual unit in word recognition. Using the probe technique (Reicher, 1969) two levels of visually similar alternatives (similar and distinct) were presented as alternatives following both single letter and word presentations. Presentations consisted of (*a*) single letters followed by two response alternatives of low visual confusability, (*b*) single letters followed by two alternatives of high visual confusability, (*c*) four-letter monosyllabic words followed by two letter alternatives of low visual confusability, and (*d*) four-letter monosyllabic words followed by two visually similar letter alternatives. For single-letter presentations, performance should be less accurate when visually similar letter alternatives are presented than when distinctive letters are presented. Similarity should also operate differently in words than in single letters if a spelling pattern or a word is the perceptual unit. On the other hand, if the letter is the perceptual unit in word recognition, similarity should have equal effects in both letter and word presentations.

TABLE 1

STIMULUS ITEMS USED IN EXPERIMENT I

Letter position tested	Stimulus material		
	Base word	Alternatives	
		Similar	Distinct
1	REAL	P	M
	BEND	R	L
	PLOW	B	G
	NAIL	M	P
	CASH	G	R
	EAST	F	C
2	SNUG	M	L
	WAIT	H	R
	AGES	C	P
	SHIN	A	P
	SHIP	K	N
	SKIM	H	W
3	PACE	G	N
	LANE	M	C
	ROBE	P	S
	CURE	B	T
	MOPE	R	L
	TOLL	I	O
4	SLAP	B	M
	STAB	R	Y
	HALE	F	L
	HEAR	P	L
	RICH	K	E
	GRIM	N	P

EXPERIMENT I

Method

Subjects. Nine undergraduate students were each paid $1.50 an hour to participate for five 1-hr. sessions. All *S*s had normal or corrected vision.

Material and apparatus. The word stimuli were 72 monosyllabic four-letter words. The 24 base words are presented in Table 1. The other 48 words are derived by replacing the letter of the base word with the letters corresponding to the similar and distinct response alternatives. The two critical letters that formed the response alternatives in the forced-choice mask were either visually similar alternatives (determined by letter overlap, Connor, 1971) or visually distinct alternatives. Either letter could be substituted to form a new word. For example, given the word LANE, similar alternatives for the third letter consisted of M and N in the mask and visually distinct alternatives consisted of C and N in the mask. The critical letters appeared equally often in the top and bottom positions. This requirement was met for each of the four letter positions. The single-letter stimuli consisted of the same letters and were presented in the same position as in word presentations. The test alternatives were presented in the same way as on word trials, including both similar and distinct pairs.

All stimuli consisted of uppercase letters typed on white index cards. The letters, typed with a Sight-Saver typewriter, subtended a visual angle of 14 min. in height and averaged about 8 min. per letter in width. A three-channel tachistoscope (Scientific Prototype, Model GA) was used to present the stimulus material. The luminance, measured by a Macbeth Illuminometer, was 16.2 ftl. for all three stimulus fields.

Procedure. On each experimental trial a fixation point was first displayed on a blank field. A ready signal from *E* preceded each trial informing *S* to initiate the trial by pressing a hand switch to immediately expose the stimulus. This was followed by the visual noise mask of blocks of typewritten overlapping SS and OS occupying the position which the stimulus had just occupied. The two response alternatives were displayed directly above the mask and above the position of the probed-for letter. Underscores were used to indicate the position of the critical letter in the stimulus. The *S*'s task was to report aloud which of the two test alternatives occurred in the stimulus presentations.

The *S*s were extensively tested, using a repeated measures design, over 5 days of the experiment. Each *S* contributed 48 observations to each combination of the major variables of the experiment: type of stimulus material (letters and words), similarity of the stimulus alternatives (similar or distinct), and four different positions of the critical letter.

On Day 1 of the experiment, *E* determined the stimulus duration at which each *S* performed at 75% accuracy. The method for determining the duration was a modified method-of-limits procedure. The test items (single letters and four-letter words) consisted of 100 items constructed similarly to those used in the experiment except that all 26 letters were used. Also, stimulus similarity was not manipulated in this set of items. The stimulus durations necessary for 75% correct were within the range of 30–75 msec. The visual noise mask containing the stimulus alternatives followed the stimulus immediately and lasted 150 msec.

The stimulus material was sorted into four blocks of 48 stimulus items with each block containing an equal number of letter and word trials. Within each block, the order of presentation of type of stimulus material, visual and distinct alternatives, and the position of the critical letter was completely random. A short rest period followed each block of trials. This procedure was maintained on the succeeding days of the experiment.

Results

Replicating previous findings (Reicher, 1969), accuracy for words (73%) was superior to that for single letters (64%), $F(1, 8) = 18.23$, $p < .05$. In contrast to our expectations, however, identification given the distinctive test alternatives did not exceed performance with similar test alternatives. This result was consistent across both letter and word presentations and all serial positions.

Discussion

The failure to find a main effect of similarity was unexpected. In contrast with previous positive effects of similarity, the present results indicate that *S*s do not seem to be able to employ partial information, i.e., visual features in the Reicher (1969) task. For example, consider the presentation of the letter D, in which the *S* sees only the curved feature. When presented with D and K as distinct alternatives he should be able to choose the correct alternative. On the other hand, this partial information would not be sufficient to discriminate the similar alternatives D and O. If *S*s were using partial information in their response selection, one would expect an advantage in discriminating distinctive alternatives. Failure to find a similarity effect suggests that *S* perceives or synthesizes the test letter or word before the alternatives are presented. In this case, if *S* did not synthesize the test item correctly, he would have to guess randomly between the two response alternatives.

If *S*s arrive at a decision about the test stimulus before the alternatives are considered, it is possible that redundancy is still available and operating to facilitate performance when letters are presented in words. During word presentations, redundancy would serve to reduce the number of alternatives, and therefore, *S* would need less visual information for correct letter identification. To illustrate the effect of redundancy, assume that *S* has read out WOR_ in Positions 1, 2, and 3 and a curved feature of the last letter ( )). From the visual information in the curved feature, the alternatives for the last position can only be D. O, or Q. However, if *S* knows the letters must spell a word, D is the only valid alternative. Accordingly, *S* synthesizes the letter configuration as WORD and is correct when presented the alternatives D and K. In the letter-alone case, given the same visual information of a curved feature, *S* also has three alternatives: D, O, or Q. However, without additional information from redundancy, *S* is equally likely to synthesize the letter as D, O, or Q. In the letter-alone case, then, *S* will perceive correctly only 33% of the time. Hence, redundancy serves to reduce the possible number of valid alternatives for word trials relative to letter trials, making an error more likely to occur on letter trials. This analysis indicates that it cannot be assumed that Reicher's (1969) paradigm eliminates the operation of redundancy in word recognition.

Synthesis of the visual stimulus before the test alternatives are considered also explains the failure to find a similarity effect. In the example above, if *S* correctly synthesizes a D beforehand he will choose the correct alternative from D and K or D and O. However, if he synthesizes an alternative not presented in the response alternatives, he will have to choose randomly from the two response alternatives. These results, therefore, do not directly test what visual features are employed in letter and word recognition.

EXPERIMENT II

Experiment II was designed as another attempt to investigate the perceptual units and visual features employed in letter and word identification. While Reicher's (1969) intent was to control for redundancy, it appears that presenting the alternatives after each presentation does not eliminate the operation of redundancy. The design of the present experiment offers a better control for redundancy while at the same time manipulating similarity. The *S* was given a fixed number of response alternatives (four) before the experimental session. Since *S* knows the valid alternatives before the test presentation, and therefore before perceptual synthesis, word presentations should not have any advantage due t redundancy. The study, therefore, pro vides a critical test of whether the word advantage effect is maintained without the contribution of redundancy.

Method

Subjects. Seven undergraduates served as *S*s to complete a requirement for the introductory psychology course at the University of Wisconsin. All *S*s had normal or corrected vision.

Materials and apparatus. Four stimulus alternatives (P, R, C, and G) were chosen from the Gibson (1969) matrix of distinctive features for graphemes. Two pairs of letters have a minimal feature difference that makes them visually similar and highly confusable, i.e., P and R. When paired with the other alternatives these letters have a minimal number of features in common. For example, P and R share four features in common, with only one feature differentiating the two. However, when P and R are paired with C or G they share no features in common and can be defined as distinctive. The letters also complete a common three-letter word when embedded in the center position. Therefore, letter trials consisted of presentation of one of the letters P, R, C, or G; the three-letter word trials consisted of presentation of either APE, ARE, ACE, or AGE. All stimuli consisted of uppercase letters typed on white index cards. The stimuli were typed so that the critical letter always appeared at *S*'s fixation point. All other methodological variables were the same as in Experiment I.

Procedure. On each experimental trial a fixation point was first displayed on a blank field. A ready signal from *E* preceded each trial informing *S* to initiate the trial by pressing a hand switch located to the left, which immediately exposed the stimulus display. Immediately after the stimulus presentation, a visual noise mask was presented for 150 msec. and covered the position of the word or letter. The *S* was to report aloud the letter which appeared during the stimulus presentation; for word trials, this meant responding with the letter in the center position.

On Day 1 of the experiment, *E* determined the duration at which each *S* performed at 75% accuracy in identifying the single letters P, R, C, and G by using the method-of-limits procedure. The range of durations over all *S*s was 20–60 msec. On Day 2, *S*s were informed of the letter and word alternatives and told that for word stimuli the center letter would be either P, R, C, or G. The stimulus material was sorted into four blocks of 56 observations with each block containing an equal number of observations at each stimulus condition. Within each block of trials the order of presentations was random. A short rest period followed each block of trials. The first 32 presentations on each day were treated as practice and were used to adjust the test stimulus duration to maintain *S*s at approximately 75% accuracy. This procedure was maintained on the last day of the experiment. Each *S* contributed 48 observations to each combination of the variables of the experiment: type of stimulus material (letters and words), similarity of stimulus alternatives (similar of distinctive), and the four different test alternatives.

Results

The results were first analyzed using percentage correct responses: the proportion of times *S*s responded with a letter given that particular letter was presented. From this data, percentage accuracy was determined from the total number of correct responses/total number of presentations. These results indicated that *S*s were more accurate during letter presentations (78%) than during word presentations (67%), F (1, 6) = 9.62, $p < .025$. This 11% difference, reflecting greater accuracy for letters alone than letters embedded in a word, eliminates the word superiority effect found in Experiment I.

The effect of visual similarity is best understood by an analysis of errors. If *S*s choose their responses randomly on error trials, they should respond with a similar alternative only 33% of the time. However, on 63% of the error trials, *S*s responded with a similar alternative, indicating they were more likely to respond with a similar alternative than with a distinctive alternative, F (1, 6) = 17.92, $p < .01$. Furthermore, this similarity effect was exactly the same on letter and word trials. Accordingly, the procedure of presenting the alternatives prior to the experimental session facilitated an effect of visual similarity while at the same time eliminating the word advantage effect.

Discussion

Experiment II demonstrated that if *S*s are given the set of stimulus alternatives before the experiment rather than after each trial, letters presented alone are better recognized than letters presented in words. A large effect of visual confusability was also shown for both letters in words and letters presented alone. Both of these results are opposite those found in Experiment I, which presented the stimulus alternatives after each trial. The results indicate that it cannot be assumed that presenting the stimulus alternatives after the stimulus trial eliminated the operation of redundancy in word recognition. Redundancy, therefore, accounts for previous findings that letters are recognized better in words than presented alone or in nonwords (Aderman & Smith, 1971; Gibson et al., 1962; Krueger, 1970; Reicher, 1969; Wheeler, 1970).

One unexpected result of Experiment II was the finding that single-letter recognition was actually better than word recognition. Although the extent of this effect is not clear at this point, *S*s reported that the third letter often interfered with identifying the middle letter. This finding suggests that lateral masking (Townsend, Taylor, & Brown, 1971) may have been operating during word presentations.

More importantly, the effect of similarity can be investigated when higher order information is eliminated. This manipulation served to emphasize visual feature analysis in identification. Under conditions of brief tachistoscopic exposures, *S*s are limited to the number of visual feature tests that can be made. Because this restriction requires an identification before all the alternatives have been eliminated, *S*s should be more likely to respond with a similar alternative than with a distinctive alternative. The present results indicated that visual similarity effects were equal in letter and word recognition, implying that the same visual features are used in letter and word identification. This indicates that the letter is the largest unit of analysis in visual recognition over which feature extraction occurs.

In summary, the experimental question of the present article was to determine the size of the perceptual unit employed in word identification. In answering this question, it was necessary to specify two sources of information which are available to the reader: visual information and nonvisual higher order information or redundancy. Words were better recognized than single letters when both these sources of information were available (Experiment I). However, the word advantage effect was eliminated when redundancy was adequately controlled (Experiment II). These findings and the fact that visual similarity effects operate similarly in letter and word recognition argue against a perceptual unit larger than a single letter.

REFERENCES

Aderman, D., & Smith, E. E. Expectancy as a determinant of functional units in perceptual recognition. *Cognitive Psychology*, 1971, **2**, 117–129.

Bruner, J. S., & O'Dowd, D. A note on the informativeness of parts of words. *Language and Speech*, 1958, **1**, 98–101.

Cattell, J. The time it takes to see and name objects. *Mind*, 1886, **11**, 63–65.

Clement, D. E., & Carpenter, J. S. Relative discriminability of visually presented letter pairs using a same–different choice-reaction time task. *Psychonomic Science*, 1970, **20**, 363–364.

Connor, J. M. Factors affecting parallel processing of visual displays. Unpublished doctoral thesis, University of Wisconsin, 1971.

Gibson, E. J. *Principles of perceptual learning and development*. New York: Appleton-Century-Crofts, 1969.

Gibson, E. J., Osser, H., Schiff, W., & Smith, J. An analysis of critical features of letters, tested by a confusion matrix. In, *A basic research program on reading.* (Final report on Cooperative Research Project No. 639) Washington, D. C.: U.S. Office of Education, 1963.

Gibson, E. J., Pick, A. D., Osser, H., & Hammond, M. The role of grapheme–phoneme correspondence in the perception of words. *American Journal of Psychology*, 1962, **75**, 554–570.

Gibson, E. J., & Yonas, A. A developmental study of the effects of visual and auditory interferences of a visual scanning task. *Psychonomic Science*, 1966, **5**, 163–164. (a)

Gibson, E. J., & Yonas, A. A developmental study of visual search behavior. *Perception & Psychophysics*, 1966, **1**, 169–171. (b)

Huey, E. B. *The psychology and pedagogy of reading*. New York: Macmillan, 1908.

Kaplan, G. A., Yonas, A., & Shurcliff, A. Visual and acoustic confusability in a visual search task. *Perception & Psychophysics*, 1966, **1**, 172–174.

Krueger, L. E. Visual comparison in a redundant display. *Cognitive Psychology*, 1970, **1**, 341–357.

Miller, G. A., Bruner, J. S., & Postman, L. Familiarity of letter sequences and tachistoscopic identification. *Journal of General Psychology*, 1954, **50**, 129–139.

Neisser, U. Visual search. *Scientific American*, 1964, **210**, 94–101.

Neisser, U. *Cognitive psychology*. New York: Appleton-Century-Crofts, 1967.

Reicher, G. M. Perceptual recognition as a function of meaningfulness of stimulus material. *Journal of Experimental Psychology*, 1969, **81**, 275–280.

Townsend, J. J., Taylor, S. G., & Brown, D. R. Lateral masking for letters with unlimited viewing time. *Perception & Psychophysics*, 1971, **10**, 375–378.

Wheeler, D. D. Processes in word recognition. *Cognitive Psychology*, 1970, **1**, 59–85.

Woodworth, R. S. *Experimental psychology*. New York: Holt, 1938.

(Received May 31, 1972)

Journal of Experimental Psychology
1973, Vol. 98, No. 1, 55–63

PROPERTIES OF HUMAN VISUAL ORIENTATION DETECTORS: A NEW APPROACH USING PATTERNED AFTERIMAGES[1]

JANETTE ATKINSON[2]

University of Cambridge, Cambridge, England

The orientation, size, shape, and spacing of two small bars affect the synchronous visibility of their afterimages. The effects found were similar for monocular and dichoptic viewing conditions, suggesting involvement of central contour detectors in the visual cortex. Synchrony was maximum for two parallel bars, lower for two bars at 90°, and lowest for two bars at 45°. For parallel bars, or bars at 90°, the afterimages synchronized even when a temporal delay was introduced between the onset of the afterimage of one bar and that of the delayed bar, indicating interaction of the bars. The visibilities of two afterimages at 45° were independent. The relative difference in orientation up to which synchrony remained at a maximum was 2° for either a vertical or horizontal "constant" bar, but 4° when the constant bar was 45°. Differences in angular selectivity for Hubel-and-Wiesel-like "hypercomplex" units are postulated to explain the results.

Many studies have shown that patterned afterimages do not fade uniformly, but with fluctuations in the visibility of the entire image or part of the image (Atkinson, 1970; Bennet-Clark & Evans, 1963; Evans, 1966; Helmholtz, 1925). There are many similarities between the fragmentation of afterimages and of optically stabilized images (Barlow, 1963; MacKinnon, Forde & Piggins, 1969). Factors such as the relative orientation, size, and locus of the parts making up the pattern seem to determine the course of fading.

MacKinnon et al. (1969), using the afterimage of an outline square as the stimulus, showed that certain combinations of lines in the pattern appeared and disappeared synchronously at a rate significantly above that expected if the lines were acting independently. The present study was carried out to determine those properties of a simple two-bar pattern that control synchronous visibility and to make quantitative measures of the effects. Two different methods of measuring synchronous visibility were used. In Experiments I, II, and III, the measure is called "synchrony," and in Experiment IV, the measure is called "correlation." A discussion of the theory behind each empirical measure precedes the description of the experiments.

EXPERIMENT I

Measurement of Synchrony

In Experiment I, the afterimage was a pattern of two small rectangular bars. The measure of visibility used was the total amount of time the afterimage was visible within the trial time of 60 sec. Three scores were obtained for each trial, these being: (*a*) The total time for which Bar A was visible whether or not Bar B was visible (t_A), the probability of Bar A being visible at any moment in the trial being P_A. In this case P_A is equal to $t_A/60$. (*b*) The total time for which the second bar (Bar B) was visible whether or not Bar A was visible (t_B), the probability of Bar B being visible at any moment of the trial being P_B, where $P_B = t_B/60$. (*c*) The total time for which both bars were visible together (t_S). The times t_A and t_B will both include t_S.

If the visibilities of the two afterimages were completely independent, then the probability of seeing both afterimages together at any moment would be the product of the probabilities of seeing each of the bars, and thus, on a 60-sec. trial the expected t_S could conceivably be taken as equal to $(P_A \times P_B \times 60)$ sec. Then the extent to which t_S exceeds $(P_A \times P_B \times 60)$ might be taken as a measure of synchrony that is present over and above chance.

---

[1] This research was supported in part by a Scientific Research Council Research Scholarship, the George Henry Lewes Studentship from the Physiological Laboratory, University of Cambridge, and the Henry George Plimmer Fellowship from Imperial College of Science and Technology, University of London.

[2] Requests for reprints should be sent to Janette Atkinson, Psychological Laboratory, University of Cambridge, Downing Street, Cambridge, England.

There are two reasons why this calculation of synchrony above chance is not appropriate in the case of two afterimages. First, the appearance of an afterimage varies systematically with viewing time. Thus the afterimages of both bars in any two-bar pattern are always visible for at least the first 5 sec., and only subsequently is the fading such that either or both may appear from time to time. There are individual differences in the length of the first period of stable synchronous visibility, but it is relatively constant for each individual. Obviously therefore, the degree of statistical independence varies throughout the trial, making inappropriate the above calculation of synchrony which assumes a constant probability for synchrony throughout the trial. A measure of synchrony must therefore be used which does not make such assumptions.

The second unsatisfactory feature of this method for calculating synchrony above chance is its failure to take into account individual differences in the absolute time for which any afterimage is visible. For example, using the above measure, synchrony above chance will be 0 sec. if t_A and t_B are 30 sec. and t_S is 15 sec. However, if the same ratios hold for another *S*, but the afterimages are seen for a shorter absolute time (t_A and $t_B = 10$ sec.; $t_S = 5$ sec.), the synchrony above chance will be 3.3 sec. This problem can be overcome by the use of the "effective trial time," i.e., the total time for which at least one bar is visible or the time for which any part of the afterimage is visible. Effective trial time, ignoring visibility of small fragments of the bars, will be equal to $t_A + t_B - t_S$.

The measure of synchrony which was used in the following three experiments is given by the ratio of two probabilities: the probability of one bar being visible when the other bar is visible and the probability of the bar being visible when anything is visible. In terms of the three measures, t_A, t_B, and t_S, made in each trial, an estimate of the probability of Bar A being visible when Bar B is visible can be taken as t_S/t_B. The probability of Bar A being visible when anything is visible can be estimated as $t_A/(t_A + t_B - t_S)$. The ratio of these two probabilities, which is called synchrony, is equal to $t_S\ (t_A + t_B - t_S)/(t_A \times t_B)$. This adopted measure of synchrony has the advantage over a simpler measure such as $t_S/(t_A \times t_B)$, in that it adjusts each *S*'s results by using the effective trial time so that comparisons can be made across individuals and it does not make any assumptions about the expected chance level of synchrony. It is used as a tool for comparing across different experimental conditions, so that the effects on synchrony of changing factors such as relative orientation and viewing conditions can be assessed.

Method

Subjects. Six *S*s were students at the University of Cambridge. They were between 17–30 yr. of age and were all naive as to the purpose of the experiments. All had vision which was normal or corrected to normal.

Apparatus and procedure. The afterimages were produced by a Braun F800 electronic flashgun. The flash duration was found to be approximately 4 msec. by displaying a photocell output on an oscilloscope. The manufacturer's specification for light output was approximately 160 joules. A diffusion filter was placed in front of the flash tube. The flash was given in a room dimly illuminated at approximately .3 cd/m^2. Patterns were produced by cutouts in a black mask placed in front of the diffusion filter. Both monocular and dichoptic viewing conditions were used. In the dichoptic viewing condition, two black shades were arranged in *S*'s line of sight between *S* and the flashgun, so that he could see only one bar in the pattern with his right eye and the other bar with only his left eye. In the dichoptic condition, any synchrony found is presumed to be due to processes involving binocular combination of information. For monocular viewing, *S* fixated, with his right eye, a white spot attached to the mask. His left eye was covered throughout the trial by an eye patch. His eyes were 28.5 cm. from the flash and horizontally aligned with it. Immediately after the flash, *S* closed his eyes and kept them closed throughout the trial time, thus maintaining a constant background luminance for viewing the afterimages.

The trial time was always 60 sec. immediately following the flash. Each *S* always had 15 min. of adaptation before the first experimental trial and 5 min. between trials. Afterimages lasted on average 2–3 min., although only the first minute was analyzed for each trial.

Certain incidental factors, known to affect the visibility of afterimages, were kept as constant as possible. The *S* was asked to fixate the spot on the command "Ready" and to avoid blinking. Orders of presentation that balanced for practice effects between different stimuli and viewing conditions were used.

The four patterns and the fixation spot used in Experiment I are shown in Figure 1. The fixation spot was positioned 2° from the nearest edge of one of the bars and horizontally aligned with the midpoint of the separation of the bars. The relative orientation of the bars was 0° or 90°. The distance separating the bars was set at 1°, 2°, 3°, or 4°. All bars were 2° × .25°.

Each *S* completed three trials on each condition, the first trial being practice. Synchrony values, calculated from the measures t_A, t_B, and t_S, were obtained for each *S* on each condition. The mean synchrony for each condition was calculated.

Results

The results for mean synchrony values are shown in Figure 2. Analysis of variance on the synchrony scores showed that: (*a*) There was a significant difference between synchrony values for different patterns, F (3, 15) = 26.9, $p < .01$. The pattern of two horizontal bars and that of two vertical bars showed higher synchrony values than patterns of two bars at 90° to one another.

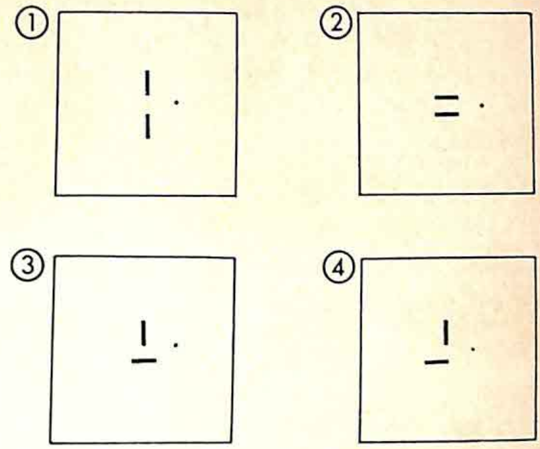


FIG. 1. The four patterns used in Experiment I. (The black dot in each figure marks the fixation point.)

(*b*) Comparing the means for each condition, there was slightly more synchrony for monocular than for dichoptic viewing, but the difference was not significant. (*c*) Synchrony was dependent on the distance separating the bars, being highest when the bars were at 1° separation, F (3, 15) = 21.8, $p < .01$.

The results of Experiment I show that relative orientation and separation affect synchrony for monocular and dichoptic viewing alike and this result implies that the variation in synchrony reflects the parameters of some process beyond the point of binocular combination of informa-

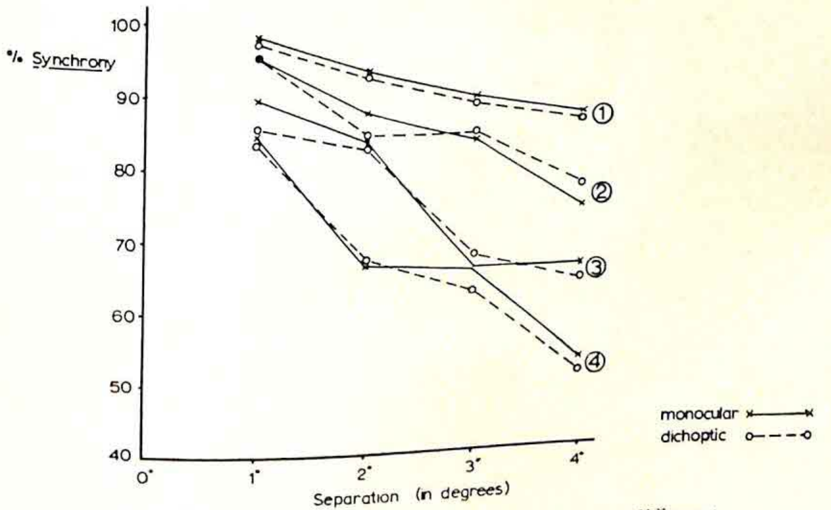


FIG. 2. Mean synchrony for different patterns at different separations (see Figure 1).

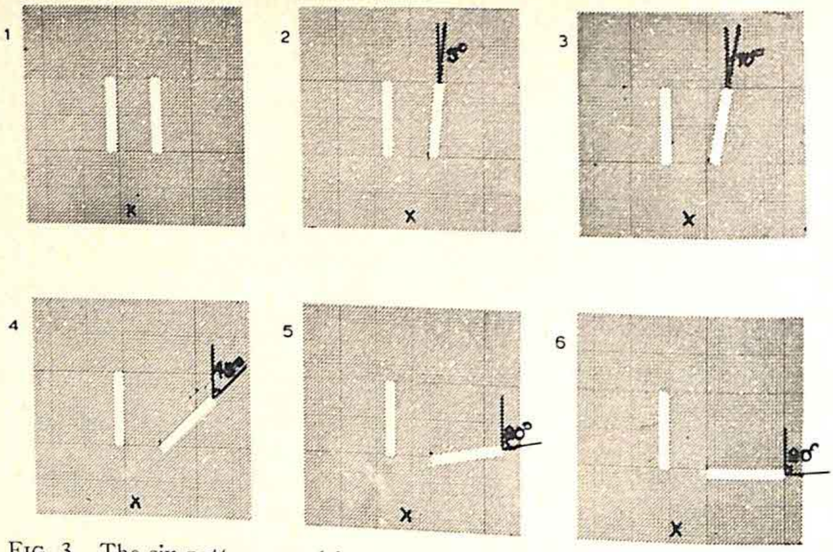


FIG. 3. The six patterns used in Experiment II (x marks the fixation spot).

tion. This means that the units mediating synchrony must be within the human visual cortex.

EXPERIMENT II

It is impossible from the results of Experiment I to know whether synchrony is low for all two-bar patterns where the bars are in different orientation, or whether synchrony varies systematically as a function of relative orientation. To investigate this, Experiment II and Experiment III were carried out.

Method

Stimuli and procedure. The basic method and
procedure were the same as in Experiment I. The
patterns used are shown in Figure 3. All bars were
2° × .25° and 1° apart. The fixation spot was ap-
proximately 2° from the base of the bars, as shown
in the figure, so that both bars were viewed para-
foveally. The six *S*s viewed the patterns mono-
cularly and completed three trials on each pattern,
the first being a practice. Mean synchrony was cal-
culated for each pattern as in Experiment I.

Results

A graph of synchrony as a function of relative orientation is shown in Figure 4. As the difference in orientation increased 0°–45°, synchrony decreased; beyond 45° there was a steady increase in synchrony up to a 90° difference in orientation. Analysis of variance on the scores for different patterns revealed a significant difference between synchrony for different degrees of relative orientation, $F\ (5,\ 25) = 18.5$, $p < .01$. Further analysis using the Tukey (*a*) test (see Winer, 1962) showed that synchrony decreased significantly for an orientation difference of 5° for two bars from maximum synchrony for two parallel bars. There was also a significant difference in synchrony for two bars at 90° to one another and two bars at 80°. There was no significant difference between relative orientations of 5° and 10°. Most of the change in synchrony occurred over the first 10° shift from the parallel bars situation.

EXPERIMENT III

Method

Stimuli and procedure. The method and procedure were essentially the same as in the previous two experiments. The patterns used in Experiment III were similar to those of Experiment II. Three different orientations of the "constant" (i.e., unchanging) bar were used, these being vertical, horizontal, and slanted at 45°. The second bar in each pattern was at 0°, 1°, 2°, 4°, 5°, and 10° to the constant bar. The bars were the same size as in previous experiments and separated by 1° in all 18 patterns. They were viewed monocularly at a position 2° parafoveal. The six *S*s (different from those in Experiment II) had five trials on each pattern, the first of the five being a practice. Mean synchrony for each pattern was calculated.

Results

The mean synchrony for each pattern is shown in Figure 5. The slope of the synchrony functions are very similar for patterns with the vertical or horizontal constant bars; synchrony did not start to decrease from maximum until there was a relative difference in orientation of 2° between the bars. For patterns with a 45° constant bar, the decline was not apparent until there was a difference of 4° in orientation. This result was confirmed by analysis of variance, $F\ (5, 25) = 34.6, p < .01$, for a difference between different degrees of orientation, and by the Tukey test for comparison of means (significant difference between 2° and 4° relative orientations for vertical and horizontal standards). There was a significant difference between 4° and 5° relative orientations for the patterns with 45° constant bars.

EXPERIMENT IV

The above results using the synchrony measure do not give any clue as to whether or not the two afterimages, which are synchronizing in visibility, are interacting with one another. Thus, if each bar had a varying time course of visible and invisible sequences, the two bars could be visible synchronously if their time courses were identical, but independent, in which case they would merely be "in phase"; alternatively, the time course of one bar could "drive" the time course of the second bar so that they appeared in phase and were not independent. In Experiment IV an attempt was made to find out whether the units mediating synchronous visibility were doing so through interaction or whether two independent units (one for each bar in the two-bar pattern) were involved.

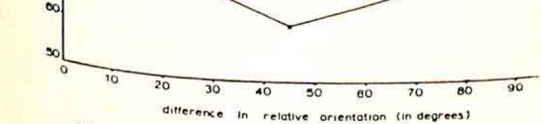


FIG. 4. Synchrony as a function of relative orientation.

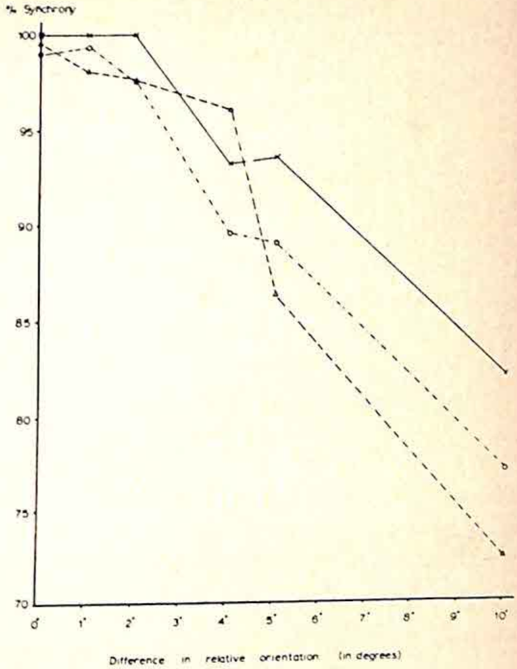


FIG. 5. Synchrony as a function of relative orientation: ×—patterns with a horizontally orientated constant bar; ○—patterns with a vertically oriented constant bar; △—patterns with a 45° oriented constant bar.

Measurement of Correlation

The two afterimages (one for each bar) were triggered by different flashes, thus introducing a delay between the onset of the first and second afterimages. If the units are independent, a delay in onset should reduce the synchrony compared with that observed for a simultaneous onset of the two images. If the units are interacting with one another to cause synchrony, there should be no reduction with the introduction of a delay.

The measurement used for measuring synchronous visibility in Experiment IV was called the correlation. It is defined as the number of seconds (rounded to the nearest second) in a given trial time for which two afterimages are in the same state (the two possible states being visible or invisible). If there was maximum "anti-synchrony" between two afterimages, then the correlation would be zero. The correlation measurement has the advantage over the synchrony measure in that it includes both states of the afterimage rather than just the visible state. However, the correlation score could give misleading results if afterimages were visible to *S* for only a very short time, as in this case the score would be spuriously inflated by the large amount of time both afterimages were invisible. The synchrony measurement has the advantage of overcoming such large individual differences by using the effective trial time rather than the actual trial time. To compare these differences between the two measurements, both correlation and synchrony were calculated in Experiment IV.

Method

Apparatus and procedure. Basically, the same apparatus and method as in previous experiments were used, except that *S*'s switches were attached to a Behavioral Research and Development four-channel event recorder. The patterns used in Experiment IV were similar to those of Experiment I. Patterns 1, 2, and 3 were the same as 1, 2, and 4 in Figure 1. Pattern 4 for Experiment IV was a vertical bar and a 45° bar (Pattern 4, Figure 3).

The interval between the onset of the afterimage of one bar and the onset of the afterimage of the other bar was set at either 0, 3, 5, 7, or 10 sec. For each flash only one bar was displayed. By means of a removable mask, *E* covered one bar for the first flash and the other bar for the second flash.

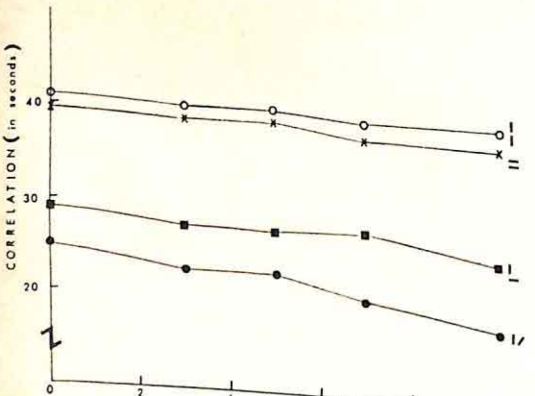


FIG. 6. Mean correlation for different patterns at different delay intervals.

From the event recording, a punched paper tape was made for each complete trial. Each tape was a list of Os for off seconds and Is for on seconds of trial time, thus representing the state of the afterimage for each second of trial time. A correlation program, used on a Modular 1 Computer, compared the two tapes for each bar of the pattern, second by second and thus gave the resulting correlation score (the total number of seconds that the two afterimages were in the same state). Synchrony for each pattern was also calculated as in previous experiments.

Each of the four *S*s recorded visibility of the afterimages for each experimental condition. The top and bottom bars were alternated as the delayed afterimage on successive trials for each *S*. Only 45 sec. of the trial time were used to calculate the correlation. This 45-sec. period started 15 sec. after the onset of the first afterimage and ran until the undelayed afterimage had a trial time of 60 sec. This procedure was used to avoid *S*'s attempting to record the visibility of the afterimage of the first bar while anticipating the second flash.

Results

Analyses of variance were carried out on both the correlation and synchrony scores for each *S*, pattern, separation, and delay interval. Both correlation and synchrony varied significantly from pattern to pattern, $F\ (3, 9) = 349.4$, $p < .01$, for correlation; $F\ (3, 9) = 173.8$, $p < .01$, for synchrony. This result confirms that of Experiment I, synchronous visibility being higher for two bars in the same orientation than for a pattern of two bars in different orientations. The mean correlation scores were 38.6 sec. for two vertical bars, 36.9 sec. for two horizontal bars, 28.4 sec. for two bars at 90°, and 23.0 sec. for two bars at 45°.

Correlation and synchrony varied significantly with the distance separating the bars, $F\ (2, 6) = 59.8$, $p < .01$, and $F\ (2, 6) = 11.7$, $p < .01$. These results also agree with those of Experiment I; the closer the bars are to one another, the greater is the synchronous visibility.

No significant overall difference was found in the correlation for different delay intervals. However, there was a significant Patterns × Delay Intervals interaction, $F\ (12, 72) = 2.8$, $p < .01$. The mean correlation for each pattern at each delay interval is shown in Figure 6. It can be seen that for patterns of two vertical or two horizontal bars, the length of the delay has a negligible effect on the correlation. There is a small decrease in correlation with increase in delay for a pattern of two bars at 90° to one another, and a larger decrease with increase in delay for the pattern of two bars at 45°. This latter result suggests that the bars in a 45° pattern are not interacting but are acting independently.

Both measures (correlation and synchrony) give very similar results. They differ logically in their treatment of the period when neither afterimage is visible, which is taken as evidence of positive correlation in the correlation score, but is excluded from the effective trial time in the synchrony measure. Thus the synchrony score might omit some significant disappearances and the correlation score might be spuriously inflated if the afterimages were only visible for a small precentage of the total trial time. The good agreement between the two scores suggests that neither of these defects seriously distorts the comparisons made between different patterns and conditions.

"*Shifted Correlation*"

A second correlation score was calculated from the original tapes and was called the shifted correlation. The shifted correlation was calculated in the same way as the original correlation score, except that instead of the same second in real time being compared for the two records, corresponding seconds in the "life" of the two afterimages were compared. Thus 1 would be added to the correlation score if both afterimages were in the same state 20 sec. after their respective initiating flashes. One can imagine the record for the second afterimage being shifted so that the starts of the two afterimages are aligned. In Experiment IV, if the visibilities of the two afterimages were independent, then the shifted correlation would be a measure of the correlation of the time courses between any two after-images and might be expected to be higher than the correlations when a delay was imposed. Calculation of the shifted correlation is not made, of course, for the zero-delay condition, because when there is no delay the shifted and original correlations are the same.

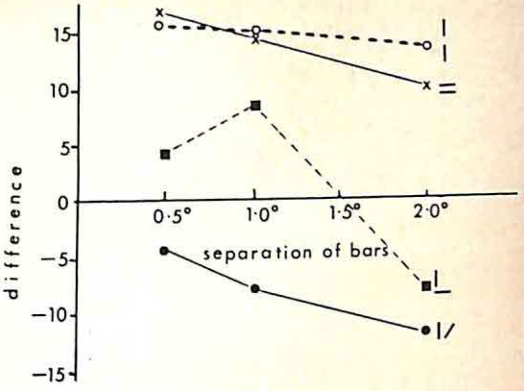


FIG. 7. The difference, in seconds, between the original and shifted correlations for different separations of the bars.

The difference between the shifted and original correlation scores was calculated for each pattern and is shown in Figure 7. Analysis of variance on these differences revealed a significant difference between different patterns, $F(3, 9) = 233.8, p < .01$, but not for different delays, $F(3, 9) = 1.6$. Figure 7 shows that the shifted correlation is higher than the original for a pattern where the bars are at 45° to one another. The original correlation is higher than the shifted correlation for a pattern of two parallel bars at all separations. The result shows clearly that for patterns in which the bars are in the same orientation, visibility of the two afterimages is not independent; one bar is visible at the same time as the second bar more frequently than such a coincidence by chance would allow. There is no evidence to suggest that synchrony in the appearance of patterns of bars at 45° to one another is anything other than independent afterimages following the same time course. For a pattern of bars at a right angle to one another, the bars seem to be interacting at small separations, i.e., 1° of separation or less.

DISCUSSION

The results have shown that synchronous visibility depends primarily upon the orientation and separation of the two bars making up the pattern. For two bars in the same orientation, correlation and synchrony are high; whereas for bars at 45° to one another, synchrony and correlation are very low. As the space between the bars increases, the correlation decreases. For patterns of parallel bars or bars at a right angle, there is some type of interaction between the bars, so that synchrony is maintained in spite of delays in onset of one of the afterimages. For bars at a right angle, this interaction only occurs if the bars are 1° or less apart; whereas for parallel bars, it is still operating at 2° of separation. For bars at 45° to one another at all separations, the bars seem to behave independently.

The results of Experiment II show that there are units sensitive to parallels or right angles, having band widths of approximately 5°–10°. This result is in good agreement with the results of Campbell and Kulikowski (1966). However, the results of Experiment III indicated that the initial decline from maximum synchrony for units sensitive to parallels depends on the orientation of the standard bar. The units are more sensitive for two horizontal or two vertical bars (tolerance 2°) than for two slanted bars (tolerance 4°). In Experiment III, synchrony appears to decline at approximately detection threshold for parallelism in afterimages, for when *S*s were asked after the experiment to give subjective judgments as to whether, in a given pattern, the bars appeared parallel or not, they reported that for patterns where the standard was vertical or horizontal the bars appeared parallel up to a difference in orientation of 2°, but for two slanted bars the bars still appeared parallel with a difference of 4°.

Several other psychophysical experiments have found similar differences in detection for parallelism. An early study measuring the threshold for parallelism showed that variability was lowest in the region of horizontal and vertical positions (Rogers, Volkman, Reese, & Kaufman, 1947). Rochlin (1955) found essentially the same result: At 0° and 90° the standard deviation was found to be 1°, whereas at 45° and 135° it was 3°. Andrews (1967a, 1967b) found selectivity higher for units tuned to horizontals and verticals than for units tuned to obliques. Campbell and Kulikowski (1966) also found that angular selectivity was higher for verticals than for obliques.

The units selectively sensitive to parallel bars of different orientations, which have been postulated to explain the results of Experiment III, seem to be similar in reaction to units previously shown to be selectively tuned to the orientation of single lines or bars. It is a common finding in psychophysical literature that vertical and horizontal orientations are better perceived than oblique orientations. The total visibility of afterimages and of optically stabilized images of vertical or horizontal single bars has been found to be greater than that of slanted bars (Atkinson, 1970; Schnidt, Fulgham, & Brown, 1971). Such results make it seem likely that the same units are involved in the orientation analysis of both single and parallel lines.

The facilitation between parallel bars or bars at a right angle, which causes increase in synchrony, must either involve specialized functional units in the visual system for detection of these two orientations or special connections between the functional units that are sensitive to single bars in particular orientations. As the above results have been shown for both monocular and dichoptic viewing, the hypothesized connections must be at the level of the cortex. Atkinson (1970) found that a single bar's unitary visibility was lowered by the addition of a second bar in an opposing orientation, this effect being called "inhibition." For a right angle pattern there is inhibition between the bars, reducing the total time each bar is visible, but at the same time high synchrony for the periods when the bars are visible (indicated by high synchrony and correlation scores). Consequently two levels of connections between units must be postulated; at one level there are interactions between detectors for single orientations, and at the second level synchrony is mediated and angles detected. A system similar to that found by Hubel and Wiesel (1965, 1968) in the cat and the monkey may be operating, whereby the outputs from "simple" cortical units feed into "complex" units which in turn feed into "hypercomplex" units. The differences in angular selectivity found in the present afterimage experiments suggest that the band widths of hypercomplex units detecting parallel lines vary depending upon whether their inputs are from complex units detecting slanted lines or units detecting single vertical or horizontal lines.

REFERENCES

Andrews, D. P. Perception of contour orientation in the central fovea: Part I. Short lines. *Vision Research*, 1967, **7**, 975–997. (a)

ANDREWS, D. P. Perception of contour orientation in the central fovea: Part II. Spatial integration. *Vision Research*, 1967, **7**, 999–1013. (b)

ATKINSON, J. A study of perceptual analysis using stabilized images. Unpublished doctoral dissertation, University of Cambridge, England, 1970.

BARLOW, H. B. Slippage of contact lenses and other artifacts in relation to fading and regeneration of supposedly stable retinal images. *Quarterly Journal of Experimental Psychology*, 1963, **15**, 36–51.

BENNET-CLARK, H. C., & EVANS, C. R. Fragmentation of patterned targets when viewed as prolonged afterimages. *Nature*, 1963, **199**, 1215.

CAMPBELL, F. W., & KULIKOWSKI, J. J. Orientation selectivity of the human visual system. *Journal of Physiology*, 1966, **187**, 437–445.

EVANS, C. R. A new approach to pattern perception. *Discovery*, 1966, **17**, 17–21.

HELMHOLTZ, H. VON. *Treatise on physiological optics.* Vol. II. (Trans. by J. P. C. Southall) Ithaca, N. Y.: Optical Society of America, 1925.

HUBEL, D. H., & WIESEL, T. N. Receptive fields and functional architecture in two non-striate visual areas (18 and 19) of the cat. *Journal of Neurophysiology*, 1965, **28**, 229–289.

HUBEL, D. H., & WIESEL, T. N. Receptive fields and functional architecture of monkey striate cortex. *Journal of Physiology*, 1968, **195**, 215–243.

MACKINNON, G. E., FORDE, J., & PIGGINS, D. J. Stabilized images, steadily fixated figures and afterimages. *Canadian Journal of Psychology*, 1969, **23**, 184–195.

ROCHLIN, A. M. The effect of tilt on the visual perception of parallelism. *American Journal of Psychology*, 1955, **68**, 223–236.

ROGERS, S., VOLKMAN, J., REESE, T. W., & KAUFMAN, E. L. Accuracy and variability of direct estimates of bearing for large displays. (Rep. No. 1-66 MHC 1) South Hadley, Mass.: Mount Holyoke College, Psychophysical Research Unit, 1947.

SCHNIDT, M. T., FULGHAM, D. D., & BROWN, D. R. Stabilized images: The search for pattern elements. *Perception & Psychophysics*, 1971, **10**, 295–300.

WINER, B. J. *Statistical principles and experimental design.* New York: McGraw-Hill, 1962.

(Received June 1, 1972)

Journal of Experimental Psychology
1973, Vol. 98, No. 1, 64–69

MEANINGFULNESS, PHONEMIC SIMILARITY, AND SENSORY MEMORY

MARGARET JEAN PETERSON,[1] CAROL E. EGER,
AND GREGORY G. BROWN

Indiana University

Three experiments studied the effects of meaningfulness, phonemic similarity, and serial position on sensory memory. Recall was directly related to meaningfulness and inversely related to phonemic similarity. Serial position effects were found when meaningfulness was varied but not when phonemic similarity was manipulated. There was some evidence of chunking when the message length was increased from six to nine letters and *S*s were encouraged to group the letters.

Recent models of memory (e.g., Atkinson & Shiffrin, 1968; Norman & Rumelhart, 1970; Reitman, 1970) have postulated three storage systems: two stores (the sensory register and the short-term store) that operate over brief intervals of time and a long-term storage system. Both experimental and physiological evidence may be cited in support of the contention that memory contains long- and brief-term components. However, it is less clear that memory over brief intervals of time involves distinguishably different systems, and some psychologists (e.g., Morton, 1970; Peterson & Johnson, 1971) have suggested that two memory stores, the sensory register and the long-term store, may be sufficient to handle the data.

One standard way of assessing the need to posit two processes is to study the effects of certain variables over the intervals of time that may differentiate the processes. If the variables have the same effect, regardless of the time intervals, then a single process is indicated; if the variables have different effects over the time intervals, multiple processes are implicated.

The present experiments tested the effects on sensory memory of meaningfulness and phonemic similarity, variables known to affect the time intervals usually identified with short-term memory (STM). Short-term memory is facilitated by increases in meaningfulness (Lindley, 1963; Peterson, Peterson, & Miller, 1961) and by the presentation of phonemically dissimilar materials (Conrad, 1967; Wickelgren, 1965) when articulation is not controlled (Peterson & Johnson, 1971). To the extent that sensory and short-term memory are affected similarly by the variables, it would be predicted that sensory memory would be directly related to the meaningfulness of the materials and inversely related to the phonemic similarity. The serial position effect also was evaluated, although serial position effects previously were demonstrated with sensory memory (Peterson, Meagher, & Ellsbury, 1970).

The basic experimental techniques used in these studies were essentially the same as the procedure described in an earlier paper (Peterson et al., 1970). A message containing six letters (Experiments I and III) or nine letters (Experiment II) was shown long enough to provide an average viewing time of 250 msec. per letter. The duration should be sufficient to permit processing without masking effects. Position probes were used to test retention. This procedure clearly involves visual (iconic) memory but may recruit auditory (echoic) and articulatory memory as well. The contribution of the various aspects of sensory memory were particularly likely to be elicited in Experiments I and II. In these experiments *S*s read aloud the letters as they appeared. The encompassing term, sensory memory, was employed to acknowledge the multiple contribution of the various sensory components.

---

[1] Requests for reprints should be sent to M. J. Peterson, Department of Psychology, Indiana University, Bloomington, Indiana 47401.

EXPERIMENT I

Experiment I tested the effects of varying the association value of six-letter messages. Each message contained two CVC (consonant–vowel–consonant) trigrams, both of which were either high or low in scaled association value. Experiment I used six letters per message to restrict the message length to one which probably was within the sensory memory domain when the items occupied adjacent spatial locations (see Neisser, 1967, chap. 2; Peterson et al., 1970) and one position was probed for each message. The total time for display of the message was 1.5 sec. and the position probe was presented immediately.

Method. Seventy-two high-association-value CVCs (90%–97%) and 72 low-association-value CVCs (30%–37%) were drawn from Archer's (1960) norms. The high-association-value CVCs (High A-V) were combined into 36 pairs to make six-letter messages with no letter overlap. The same procedure was followed to prepare 36 six-letter messages containing two low-association-value CVCs (Low A-V) with no letter overlap. In addition to selecting the trigrams from restricted ranges of association value, transition probabilities were computed, using the frequencies tabulated by Underwood and Schulz (1960). For each trigram, the average transition probability of the likelihood that the second letter would follow the first, that the third letter would follow the second, and that the third letter would follow the joint appearance of the first and second letters was computed separately. These probabilities for the first trigrams with high association values were .13, .03, and .05, respectively, and the corresponding probabilities for the first trigrams with low association values were .09, .03, and .03. The values for the second trigrams, High A-V, were .14, .06, .06; and for the second trigrams, Low A-V, were .10, .03, and .02. In general, the transition probabilities were higher for the High A-V messages than for the Low A-V messages.

The High A-V and Low A-V messages were randomized to blocks of 12 so that six High A-V and six Low A-V messages were shown in each block of 12 trials, and each of the six positions was probed for both High A-V and Low A-V messages.

The list of 72 messages was printed by an IBM printer on computer paper, prefaced by three practice messages. Six copies of the printout were prepared so that each letter of each message could be probed over the six lists (printouts). The target or probe position was indicated by an arrow that appeared as the paper advanced. One blank space separated the probe from the next message to give *S* a 1.5-sec. rest.

Each message was displayed on a device constructed from an IBM paper puller for 1.5 sec. or for an average display time of 250 msec. per letter. Three seconds were allowed for recall.

The *S*s were 24 introductory psychology students at Indiana University who participated as one alternative to fulfilling a course requirement. The *S*s were assigned randomly to four replications of the block of six lists.

The *S*s read aloud the message as soon as it appeared and then tried to recall the letter that formerly occupied the position of the arrow.

Results and discussion. The *S*s recalled 92% of the letters from High A-V messages and 85% of the letters from Low A-V messages, F (1, 1933) $= 21.28$, $p < .01$, indicating that sensory memory is influenced by the meaningfulness of the materials. The data also showed a typical serial position effect, F (5, 1633) $= 9.36$, $p < .01$. The entire serial position curve for the High A-V messages was higher than the curve for the Low A-V messages, as shown in Table 1, and the interaction was not significant, F (5, 1633) $= 1.57$, $p > .05$.

The percentage recalled was lower on the first block of 12 trials (80%) than on the others which increased from the second block (89%) to the sixth block (93%), F (5, 1633) $= 7.11$, $p < .01$. A Tukey test applied to the mean recall for the blocks showed that recall was significantly poorer, at the .01 level, on the first block than on the other blocks and that the other blocks did not differ.

These results demonstrate that sensory memory, like STM, is directly related to meaningfulness and shows reliable serial position effects when a single position is probed on each trial to reduce output interference and when the total number of juxtaposed letters probably is within the capacity of sensory memory.

In Experiment I, meaningfulness was varied by using CVCs of high or low association value, and, although *S* did not recall the three letters of the CVCs as a unit, more letters were recalled from the high-association-value trigrams. Experiment II determined whether *S* could retain as many as nine letters if he was encouraged to treat the individual trigrams as units or chunks. Murdock (1961) suggested that chunking was responsible for the facilita-

TABLE 1

Proportionate Recall for Each Serial Position of Messages in Experiments I, II, and III

Message	Serial position									
	1	2	3	4	5	6	7	8	9	Overall
Experiment I										
High A-V	.96	.92	.91	.87	.90	.97				.92
Low A-V	.92	.83	.78	.76	.88	.94				.85
Experiment II										
High A-V	.91	.80	.86	.96	.82	.81	.97	.96	.98	.88
Low A-V	.77	.75	.74	.72	.76	.78	.89	.89	.90	.80
Experiment III										
Phonemically similar	.74	.79	.74	.75	.75	.73				.75
Phonemically dissimilar	.78	.79	.79	.79	.83	.81				.80

Note. A-V means association value.

tion of STM when familiar words were used.

The possible facilitation by chunking was particularly interesting in view of the limited capacity that has been ascribed to sensory memory for materials occupying adjacent locations. Mackworth (1962) suggested that from three to four digits could be read from visual sensory memory. Peterson et al. (1970) set the limit at six to seven digits. If *S* does treat the trigram as a unit, then recall for all nine letters should be high. Furthermore, while the recall of the trigrams might show a serial position effect, the retention of the letters within a trigram should not differ appreciably.

If *S* does not chunk and the capacity of sensory memory is limited to six or seven letters under conditions of these experiments, then the probability of recalling the first few letters should be determined by the likelihood that these items have been processed into a longer term store. Neisser (1967) posited that information may be available in visual sensory memory for only a few seconds, unless the information is recoded into a more permanent memory. The rate of presentation of the materials in these experiments is so rapid that processing of the early material into a more permanent store might have been difficult. Hence, it was assumed that recall of the first three letters would be poorer in Experiment II than in Experiment I, unless chunking occurred for all three trigrams.

Experiment II

Experiment II showed three trigrams or nine letters per message. To encourage *S*s to chunk the letters, they were informed that they would be asked to recall the first three letters, the middle three letters, or the last three letters. If *S*s treated the trigrams as units, it was expected that the recall curves would show serial position effects between, but not within, the trigrams. On the other hand, if sensory memory encompasses only about six or seven adjacent letters, a smooth serial position curve over the nine letters should be found reflecting sensory memory for the latter letters and a longer term store for the former letters. In addition, the performance on the first of the three trigrams of Experiment II would be poorer than the performance on the first of the two trigrams of Experiment I.

Method. Thirty-six new trigrams with association values between 90%–97% and 36 new trigrams with association values between 30%–37% were taken from Archer's (1960) norms. The new High A-V trigrams were combined with the High A-V messages from Experiment I. The same procedure was followed for the Low A-V messages. Each message contained nine different letters. The transition probabilities for the first two trigrams were the values given for Experiment I. For the third trigram, the probabilities that the second letter would follow the first were .10 and .08 for the High A-V and Low A-V trigrams; the corresponding probabilities that the third letter would follow the second letter were .04 and .03, and that the third letter would follow the combination of the first two letters were .06 and .02.

The 72 messages were randomized to the six cells of a block which varied high- and low-association-value messages with the position of the CVC to be probed (first, second, or third CVC of the message). Three copies of this randomization were printed on computer paper, permitting each CVC of each message to be tested once over the three lists. A second randomization was prepared, and three copies of this list were generated. Again, each CVC of each message was tested over the three lists. The same three practice messages were used for each of the six lists.

Each message was presented for 2.25 sec., providing an average display time of 250 msec. per letter. Three probe arrows indicating that *S* was to recall the three letters of the first, the second, or the third trigram were shown for 4.5 sec. and then *S* had a 2.25-sec. rest interval before the next message was displayed.

The *S*s were 24 different students from introductory psychology courses at Indiana University. They were assigned randomly to four replications of the block of six lists. As in Experiment I, *S* was instructed to read the letters aloud.

Results and discussion. Experiment II also demonstrated that the percentage of letters recalled was higher for the High A-V messages (88%) than for the Low A-V messages (80%), F (1, 4945) = 74.88, $p < .01$, and a serial position effect was found overall, F (8, 4945) = 22.88, $p < .01$. None of the interactions was reliable.

The serial position effect was examined further for evidence of chunking. It was assumed that if *S*s treated the letters of each trigram as a unit, the recall of the individual letters within each trigram would not differ. A Tukey test was used to make these comparisons. In general, *S*s recalled the letters within each trigram about equally often, suggesting that *S*s did treat the trigrams as units. The one exception occurred for the first letter of the first trigram. Recall for this letter was reliably higher than the recall for the second and third letters of the trigram. When the average recall for the three trigrams was compared, a decided recency effect emerged. The recall for the individual letters of the third trigram exceeded the recall of any of the preceding letters, and the recall scores for the letters in Serial Positions 2–6 did not differ.

The next question of interest dealt with whether chunking resulted in recall of the nine-letter messages of Experiment II that was equivalent to the recall of the six-letter messages in Experiment I. Hence, the recall of the trigrams in the two experiments was compared.

The recall for the last two letters of the third trigram of Experiment II (Positions 8–9) did not differ from recall of the last two letters of the last trigram of Experiment I (Positions 5–6). However, recall of the first letter of the last trigram in Experiment II (Position 7) was higher than recall for the first letter of the last trigram of Experiment I (Position 4), providing additional evidence for chunking in Experiment II. On the other hand, recall for the letters of the first two trigrams of Experiment II was lower than the recall of the first trigram of Experiment I, F (1, 46) = 10.49 and F (1, 46) = 12.70, respectively, both ps $< .01$. The poorer performance in Experiment II may have been produced by output interference or by the necessity to recode early information into a more permanent memory store. The conditions of Experiment II favored chunking, but whatever chunking occurred did not reduce the memory load for the first two trigrams of Experiment II to a level comparable to that of the first trigram in Experiment I.

Experiment III

The third variable examined the effects of phonemic confusions in sensory memory. Short-term memory is adversely affected by phonemic similarity unless articulation is controlled (Peterson & Johnson, 1971). Experiment III sought to determine if sensory memory is similarly affected.

Method. Each *S* saw 144 six-letter messages. Seventy-two messages were prepared using the phonemically similar letters V, Z, T, C, G, and P, and 72 messages contained the phonemically dissimilar letters M, R, W, D, K, and F. The messages were constructed by randomly ordering the letters of the appropriate type. Transition probabilities were computed, using the method described for Experiment I. The probabilities of the second letter given the first, of the third letter given the second, and of the third letter given the two preceding letters were about the same for the dissimilar and similar trigrams. The respective values for the first trigrams were .01, .01, and .03 for the dissimilar; .01, .02, and .03 for the similar. The values for the second trigrams were .02, .01, and .03 for the dissimilar; .02, .03, and .03 for the similar. These probabilities were lower for Experiment III than for the High A-V messages of Experiments I and II and were either lower than or equal to the transition probabilities for the Low A-V messages of Experiments I and II.

The messages were then randomly assigned to blocks of 12 trials with six phonemically similar and six phonemically dissimilar messages per block. Each of the six serial positions was probed once in a block for each type of message. Six lists were prepared so that each serial position within each message was tested over the six trials.

These lists were printed on computer paper following three practice messages. The messages were shown at a 1.5-sec. rate with 3 sec. allowed for recall. The single position probed during recall was indicated by an arrow. The trials were separated by a 1.5-sec. rest interval and longer rest intervals followed every 36 trials. The *S*s did not read the letters aloud in Experiment III, because reading the letters would have deliberately introduced auditory and articulatory cues. In Experiment III, *S*s wrote their responses on prepared response sheets.

The *S*s were 24 new students from the same source. They were assigned randomly to four replications of the block of six lists.

Results and discussion. The *S*s recalled 75% of the letters from the phonemically similar messages and 80% of the letters from the phonemically dissimilar messages, $F\ (1, 2024) = 12.47,\ p < .01$. Thus, phonemic similarity appears to be inversely related to performance in both sensory and short-term memory. There was no evidence of a serial position effect ($F < 1$), nor did the serial position variable interact with phonemic similarity. Performance did tend to improve over the 12 blocks, $F\ (11, 2024) = 4.31,\ p < .01$.

Experiment III produced poorer recall than either Experiment I or Experiment II, probably because the transition probabilities of the second or third letter of each trigram conditionalized upon the previous letters were lower for Experiment III than for the other experiments. This observation would support the contention that meaningfulness, as measured by transition probabilities, does influence sensory memory. However, the transition probabilities were almost identical for the phonemically similar and dissimilar messages and thus could not be invoked to explain the differential recall shown for these two types of messages in Experiment III.

General Discussion

The evidence of these three experiments suggests that similar effects of meaningfulness and phonemic similarity are found over the time intervals usually associated with sensory memory and short-term memory. It would be more parsimonious to postulate one, rather than two, memory stores operating over brief intervals of time.

However, it must be noted that Wickelgren (1965) found evidence of associative relationships in short-term memory while Peterson et al. (1970) found little evidence of associative relationships in sensory memory. This discrepancy may have occurred because Wickelgren presented nine letters and requested complete recall. Thus, it is likely that Wickelgren's procedure tapped a longer term store. The Peterson et al. study used probes similar to the procedures employed in the present experiment. In another experiment designed to investigate "very-short-term" visual sensory memory, Wickelgren and Whitman (1970) did not observe associative responding.

In summary, the evidence obtained by the three experiments mirrored typical findings in the short-term-memory literature, thus providing support for the contention that a single brief memory store may be posited to cover what has been identified as the "sensory register" and at least the early portions of the short-term store. The portions of the short-term store which span longer time intervals may involve verbal encoding or some other transformation. Such transformations may occur with the present method of probing "sensory memory" despite the restricted presentation time.

REFERENCES

Archer, E. J. Re-evaluation of the meaningfulness of all possible CVC trigrams. *Psychological Monographs*, 1960, **74** (10, Whole No. 497).

ATKINSON, R. C., & SHIFFRIN, R. M. Human memory: A proposed system and its control processes. In K. W. Spence & J. T. Spence (Eds.), *The psychology of learning and motivation: Advances in research and theory*. Vol. II. New York: Academic Press, 1968.

CONRAD, R. Interference or decay over short retention intervals. *Journal of Verbal Learning and Verbal Behavior*, 1967, **6**, 49–54.

LINDLEY, R. H. Effects of controlled coding cues in short-term memory. *Journal of Experimental Psychology*, 1963, **66**, 580–587.

MACKWORTH, J. F. The visual image and the memory trace. *Canadian Journal of Psychology*, 1962, **16**, 55–59.

MORTON, J. A functional model for memory. In D. A. Norman (Ed.), *Models of human memory*. New York: Academic Press, 1970.

MURDOCK, B. B., JR. The retention of individual items. *Journal of Experimental Psychology*, 1961, **62**, 618–625.

NEISSER, U. *Cognitive psychology*. New York: Appleton-Century-Crofts, 1967.

NORMAN, D. A., & RUMELHART, D. E. A system for perception and memory. In D. A. Norman (Ed.), *Models of human memory*. New York: Academic Press, 1970.

PETERSON, L. R., & JOHNSON, S. T. Some effects of minimizing articulation on short-term retention. *Journal of Verbal Learning and Verbal Behavior*, 1971, **10**, 346–354.

PETERSON, L. R., PETERSON, M. J., & MILLER, A. Short-term retention and meaningfulness. *Canadian Journal of Psychology*, 1961, **15**, 143–147.

PETERSON, M. J., MEAGHER, R. B., JR., & ELLSBURY, S. W. Repetition effects in sensory memory. *Journal of Experimental Psychology*, 1970, **84**, 15–23.

REITMAN, J. S. Computer simulation of an information-processing model of short-term memory. In D. A. Norman (Ed.), *Models of human memory*. New York: Academic Press, 1970.

UNDERWOOD, B. J., & SCHULZ, R. W. *Meaningfulness and verbal learning*. Philadelphia: Lippincott, 1960.

WICKELGREN, W. A. Acoustic similarity and retroactive interference in short-term memory. *Journal of Verbal Learning and Verbal Behavior*, 1965, **4**, 53–61.

WICKELGREN, W. A., & WHITMAN, P. T. Visual very-short-term memory is nonassociative. *Journal of Experimental Psychology*, 1970, **84**, 277–281.

(Received June 3, 1972)

Journal of Experimental Psychology
1973, Vol. 98, No. 1, 70–78

ENCODING EFFECTS OF RESPONSE BELONGINGNESS AND STIMULUS MEANINGFULNESS ON RECOGNITION MEMORY OF TRIGRAM STIMULI [1]

HENRY C. ELLIS [2] AND E. CHANDLER SHUMATE

University of New Mexico

This experiment investigated the role of stimulus meaningfulness (M) and response representativeness on the temporal course of stimulus recognition and response recall of verbal items. The *S*s learned to associate representative or nonrepresentative common words with high- or low-M trigram doublets. After delay intervals of 0, 2, 7, 14, or 28 days, each *S* was given a forced-choice stimulus-recognition test and free- and aided-recall tests for the response terms. Response representativeness produced superior recognition of the low-M stimuli, whereas high-M stimuli were recognized equally well regardless of response context, supporting a conceptual coding interpretation of context effects in recognition. Stimulus-recognition functions were invariant over time, whereas response recall, both free and aided, declined systematically across delay intervals. Moreover, the proportions of correct recalls, given correct stimulus recognition, decreased over time, which supported a selection-encoding interpretation of the locus of response effects.

This experiment focuses on three related issues. The first concerns the manner by which responses affect the encoding of stimuli during the learning of associative tasks. The specific problem concerns whether stimulus recognition of verbal items following associative learning is dependent upon properties or features of the associated responses. While there is no doubt that human *S*s can form interactive relationships between stimuli and responses based upon imagery or other mnemonic processes, the question is: Does the formation of an association between stimuli and responses affect the encoding of stimuli as reflected in a stimulus-recognition test?

The limited data pertinent to this question appear on the surface to be contradictory. In several studies (e.g., Daniel & Ellis, 1972; Ellis, 1968, Experiments I & II; Ellis & Daniel, 1971, Experiments I & II; Ellis & Muller, 1964, Experiment I), positive effects on stimulus recognition as a result of associative training have been reported. These studies employed visual form stimuli (random shapes) varying in such properties as stimulus meaningfulness (M), codability, and physical complexity. In contrast, studies employing trigram stimuli (e.g., Martin, 1967) have found that stimulus recognition is unaffected by the association of responses. Similarly, Bower (1970) has examined this issue in the context of imagery research and found that recognition of nouns employed as stimulus terms was essentially independent of various associative training procedures. Since these studies differ in a number of features including type of stimulus items, type of responses, and stimulus meaningfulness, a resolution of these differences seems important. This experiment shows that the encoding of verbal stimuli, as measured by stimulus recognition, is dependent upon properties of the responses used during associative training, and that only low-M stimuli benefit from such effects.

A formulation of the conditions under which responses facilitate the encoding of stimuli has been outlined by Ellis and Muller (1964) who proposed that associating a "representative" verbal response with a stimulus enables encoding of that stimulus and hence facilitates subsequent recognition *when* the response allows *S* to relate the stimulus to some available concept for which it is representative. They assumed that the representative response effectively instructs *S* as to what information in the stimulus he may easily encode. The most important assumption in this account, which Ellis (1972) subsequently called a *conceptual coding hypothesis*, was that the effect of associating verbal responses to stimuli is dependent upon the initial *codability* of the stimuli, with the *relative* effects of associative training being greater with the less codable stimuli. Codability is used here as a generic term referring to the class of stimulus properties (e.g., physical, associative, linguistic) which control the ease with which stimulus items may be placed in memory. More generally, the hypothesis assumes that any mnemonic, whether it involves imagery instructions or response terms having properties likely to facilitate the linkage between two items, will also produce its greatest relative effect on stimulus encoding with stimuli that are "impoverished," fractionated, or initially difficult to encode.

---

[1] This research was supported by National Science Foundation Grant GB-27413X to the first author, and by a National Defense Education Act graduate fellowship to the second author. This experiment was initiated while the senior author was Visiting Professor of Psychology at the University of California, Berkeley. The support of Leo Postman and the Institute of Human Learning staff is gratefully acknowledged.

[2] Requests for reprints should be sent to Henry C. Ellis, Department of Psychology, University of New Mexico, Albuquerque, New Mexico 87106.

A second issue concerns the role of stimulus M in recognition of verbal materials. Although it is known that stimulus M can facilitate recognition of visual shapes, this effect depends in a rather complex way upon the index of meaningfulness employed (e.g., Ellis, Muller, & Tosti, 1966; Feuge & Ellis, 1969) and upon stimulus complexity as well (Ellis, 1968). Similarly, Martin and Melton (1970) have reported positive effects of M on trigram recognition; however, their use of a high-threshold model for correcting single-stimulus recognition data raises some question about the generality of their findings. Therefore, this experiment provides further evidence regarding stimulus-M effects with a forced-choice procedure, and shows that stimulus M is positively related to recognition performance.

Finally, this experiment provides additional data on the *locus* of response effects on stimulus recognition. If associating responses to verbal stimuli does facilitate stimulus recognition of verbal items, as is the case with visual pattern stimuli (e.g., Ellis, 1968; Ellis & Daniel, 1971), then one issue is the locus of this effect. Attempts to account for this effect may be classified into two categories, one focusing on the role of the associated response at the time of recognition testing (e.g., Price & Slive, 1970) and the other focusing on the role of the response during associative training (Ellis, 1968; Ellis & Daniel, 1971). The first type of account emphasizes the importance of the response as a retrieval cue, whereas the second account emphasizes the processes of stimulus selection and encoding. Although Ellis and Daniel (1971) and Daniel and Ellis (1972) obtained strong evidence for an encoding rather than a retrieval interpretation, this experiment provides additional evidence on this question when the stimuli are verbal items rather than visual patterns. These two interpretations are evaluated by comparing the temporal course of stimulus recognition and response recall and by examining the temporal course of recall given stimulus recognition. If response recall is forgotten at a rate faster than stimulus recognition, and if the probability of response recall given stimulus recognition decreased over time, then the evidence would be against the view that stimulus recognition is mediated by the response serving as a retrieval cue, because stimulus recognition will occur even when the associated response mate is not accessible.

METHOD

Experimental design. The *S*s were 200 volunteer students from the University of New Mexico. Ten *S*s were assigned at random to 1 of 20 conditions of the experiment. Two levels of stimulus meaningfulness (high and low), two kinds of responses (representative and nonrepresentative), and five retention intervals (zero delay, 2 days, 7 days, 14 days, and 28 days) were combined in a 2 × 2 × 5 factorial design.

The experiment was conducted in two stages: During pretraining all *S*s were given eight paired-associate (PA) anticipation trials to learn a list of eight pairs. The stimulus terms were trigram doublets of either high or low M and the response terms (labels) were common words which were either representative or nonrepresentative of the

TABLE 1

LISTS OF TRIGRAM DOUBLETS AND CORRESPONDING REPRESENTATIVE AND NONREPRESENTATIVE RESPONSES

Stimuli	Responses	
	Representative	Nonrepresentative
High meaningfulness		
BANCOW	BACON	INFORM
CUPNET	COVET	PREFER
RIBHUT	RABBIT	FENDER
LAWSIR	LOSER	HOTEL
RUMWIG	RUMMAGE	GRAVEL
MEDVAN	MAIDEN	UMPIRE
SUMKAY	SUNDAY	OATMEAL
DENGIT	DINGY	TOKEN
Low meaningfulness		
NEJWOQ	NETWORK	SIGNAL
YIWFEP	YIPPIE	LEVER
TIJFEQ	TIGER	ANGEL
RUVCIJ	RUBBISH	DIVER
BUVGEJ	BUDGET	ENJOY
QUJGEC	QUEBEC	FRAGILE
JUQXIH	JUSTICE	WILLOW
MOJZUV	MOTIVE	CONFORM

stimuli. Following pretraining, *Ss* were given three separate memory tests.

Items. Four separate PA lists were used, combining factorially two levels of stimulus M and two levels of response representativeness. The stimulus terms were doublets (pairs) formed by combining either two high- or two low-M CVC (consonant–vowel–consonant) trigrams taken from Archer's (1960) list. Doublets rather than single trigrams were used in order to require *S* to encode more information and to allow for convenient construction of four different distractor items (for the recognition test) in accord with a single construction rule. The average association values for the high- and low-M CVC doublets were 100% and 12%, respectively.

The response terms were common words, equated for frequency of occurrence between lists in accordance with the Thorndike-Lorge norms. Response representativeness (belongingness) was determined by agreement among six judges as to whether or not a particular common word selected resembled its stimulus mate. Judges were asked to rate each response as being representative, in the sense of being descriptive, suggestive, or relevant to the stimulus. A nonrepresentative response was judged as bearing little, if any, relationship to its corresponding stimulus mate. These procedures are comparable to those employed by Ellis (1968) in studies of visual shape recognition. In the case of verbal stimuli, these operations basically mean selection of response terms (representative) which are formally and/or acoustically similar to the stimuli. Although the defining operations of response representativeness clearly imply a multidimensional concept, all that is necessary for the present experimental purposes is the selection of two sets of pairs for associative learning: one set which is relatively easy to associate and one which is relatively difficult. Table 1 presents the lists of stimuli and responses used in the experiment.

Pretraining. All *Ss* were given eight PA anticipation trials to learn the stimulus–response pairs. Each stimulus was exposed for a 4-sec. period consisting of a 2-sec. anticipation interval and a 2-sec. simultaneous presentation with the response. The intertrial interval was 4 sec. The *Ss* were tested individually and the materials were presented using a Stowe memory drum.

Memory tests. Following pretraining, *Ss* were tested for retention at various intervals of delay. The *Ss* in the zero-delay condition were tested immediately and the remaining *Ss* returned to the laboratory for testing after delay intervals of 2, 7, 14, and 28 days. Three memory tests were given to each *S* in the following order: (*a*) *Free recall* of the response terms was obtained by having *Ss* write as many of the responses as they could recall on a sheet of paper. (*b*) A *forced-choice stimulus-recognition* test required *Ss* to select one of five highly similar stimuli as being an item seen during pretraining. The five stimuli (an old item and four distractors) were typed in a single row on a white 5 × 8 card. The distractors differed from the old item and each other by a substitution of only one letter at one of the four middle positions in the doublet. For example, the prototype BANCOW appeared with the distractors BENCOW, BAZCOW, BANHOW, and BANCUW. Given the similarity of the distractors, the task requires that accurate recognition be based upon an encoding containing most of the letters of the training items. Moreover, an encoding based upon simply the first or last letter or both is inadequate for correct recognition since these letters are completely redundant in the test. The *Ss* simply pointed to the stimulus they thought to be the one they had seen during pretraining. (*c*) *Aided-recall* measures were obtained by presenting the eight training stimuli to *Ss* on a sheet of paper and instructing them to write the corresponding response terms.[3]

RESULTS

PA learning. Table 2 shows the results of original PA learning for each condition

---

[3] The use of three memory measures does not introduce a source of confounding in this study. First, Ellis and Daniel (1971) found that stimulus recognition was unaffected by administration of a prior free-recall test when compared with conditions in which no free recall of responses was required. Second, the stimulus-recognition test should not affect the aided recall of responses since the correct stimuli are presented in the latter test.

TABLE 2

Mean Correct Anticipations on Last Learning Trial

Condition		Retention interval (in days)				
Stimuli	Responses	0	2	7	14	28
High meaningfulness	Representative	8.0	8.0	8.0	8.0	8.0
	Nonrepresentative	6.0	6.2	5.4	5.1	5.0
Low meaningfulness	Representative	8.0	8.0	8.0	8.0	8.0
	Nonrepresentative	3.8	4.2	3.0	4.8	4.7

of the experiment. Terminal performance was positively correlated with response representativeness, regardless of the meaningfulness of the stimulus terms, when the response terms were representative of the stimuli. Conversely, when the response terms were not representative, performance was poorer and was correlated with stimulus M. This latter interpretation is supported by a significant Stimulus M × Response Representativeness interaction, F (1, 180) = 11.55, $p < .001$. In addition, a nonsignificant delay effect, F (4, 180) = .61, $p > .05$, indicates that performance of a given experimental condition, for the various delay intervals, while showing some variability, was essentially comparable at the end of PA training. These data are of little interest since the data of primary interest were the scores on the recognition test.

Stimulus recognition. The data of principal interest are the stimulus-recognition results portrayed in Figure 1. This figure is a plot of the mean number of correct stimulus recognitions (calling an old stimulus "old") for the conditions of the experiment. Since this was a five-item forced-choice task, chance performance was 1.6 recognitions (20%) and performance was substantially above chance for all experimental conditions.

Inspection of Figure 1 reveals no evidence of a loss in recognition performance over the delay intervals employed. Although small fluctuations in recognition perfor-

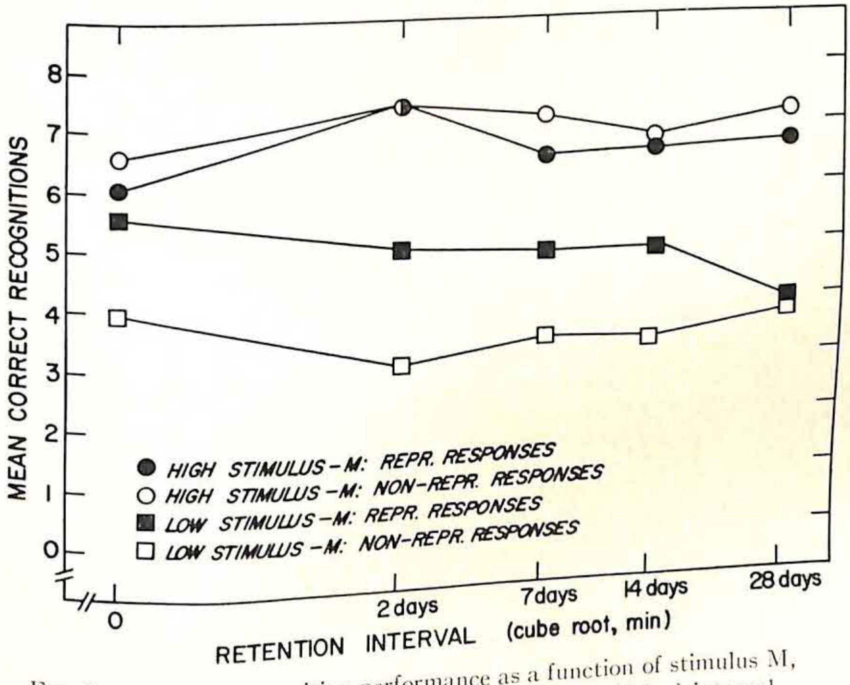


Fig. 1. Stimulus-recognition performance as a function of stimulus M, response representativeness, and length of retention (delay) interval.

mance appeared, these fluctuations were neither systematic nor reliable as reflected in the nonsignificant delay effect, F (4, 180) $= .25$, $p > .05$. Moreover, the delay variable did not interact with either stimulus M or response representativeness, nor did it contribute to a significant Delay × Stimulus M × Response Representativeness interaction (all ps $> .05$), thus attesting to the temporal invariance of the stimulus-recognition functions.

The powerful facilitating effects of stimulus M on stimulus recognition are evident with *S*s correctly recognizing an average of 86% of the high-M items and only 53% of the low-M items. This difference was highly reliable, F (1, 180) $= 170.28$, $p < .001$. Similarly, the effect of response representativeness on stimulus recognition is evident with greater recognition of trigram stimuli which were associated with representative, in contrast to nonrepresentative, responses. Although this effect was not pronounced it was reliable, F (1, 180) $= 5.17$, $p < .025$.

The importance of response representativeness, however, as a factor facilitating the encoding of a stimulus is most vividly portrayed in the interaction between stimulus M and response representativeness. Figure 1 clearly shows that associating a representative response to a stimulus obviously aids the encoding of that stimulus, as estimated by stimulus recognition, when the stimuli are low M but *not* when they are high M. The Stimulus M × Response Representativeness interaction was reliable, F (1, 180) $= 17.23$, $p < .001$.

It should be noted that differences in the degree of original learning of the PA task do not, of course, account for differences in stimulus-recognition performance. Moreover, this issue is not directly pertinent in this study since the focus is on stimulus recognition.

Free recall. The results of the free-recall test showed that the mean percentage of recall of responses for the five delay conditions, averaged across stimulus M and response representativeness, was 88%, 66%, 55%, 39%, and 27% for the 0-, 2-, 7-, 14-, and 28-day delays, respectively. Interest in these data is not in the well substantiated loss in retention over time, which is quite reliable here, F (4, 180) $= 60.45$, $p < .001$, but in the fact that systematic forgetting

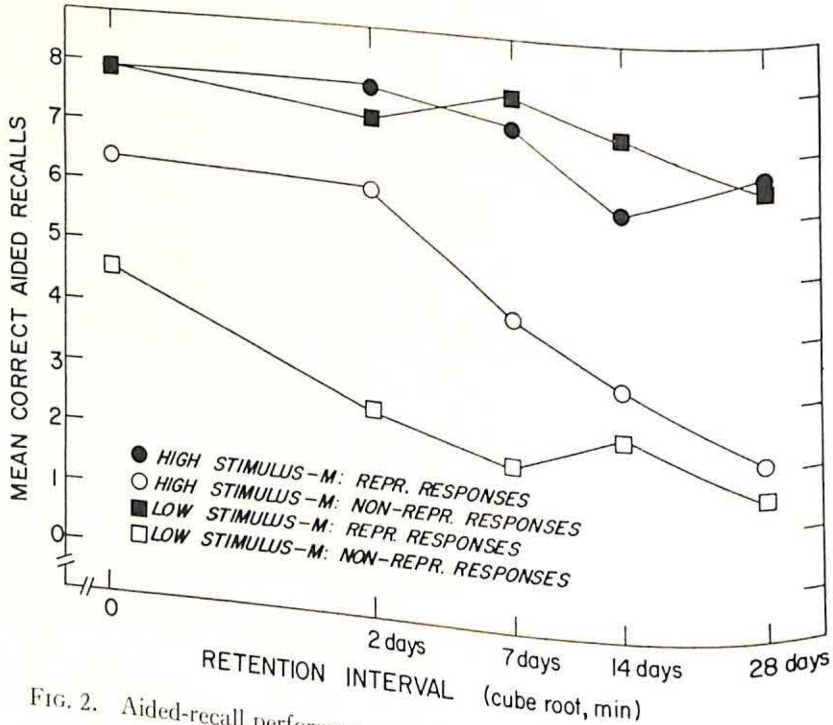


FIG. 2. Aided-recall performance as a function of stimulus M, response representativeness, and length of retention (delay) interval.

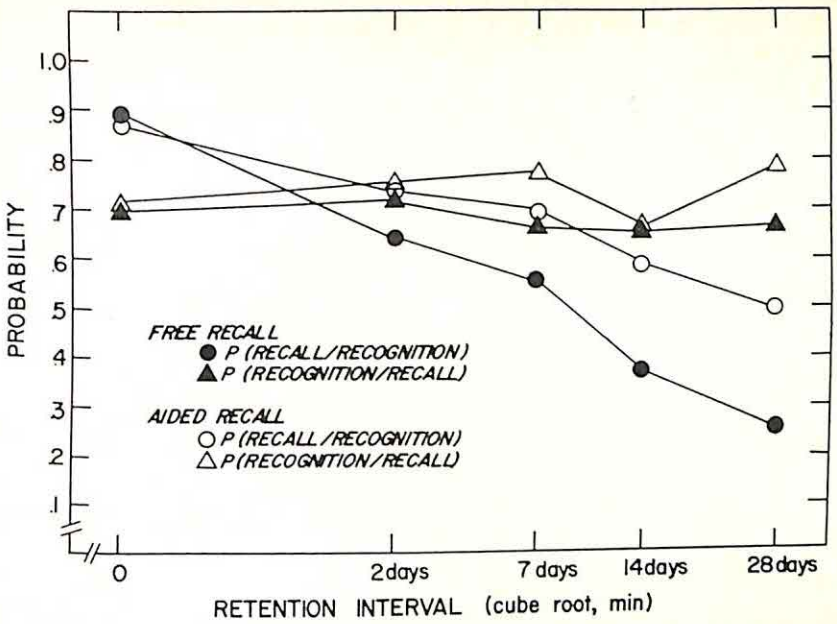


FIG. 3. Conditional probability data as a function of length of retention (delay) interval.

of the responses, given stable stimulus-recognition performance over time, makes it evident that free recall of the responses does not serve as a reliable retrieval cue in mediating stimulus-recognition performance. Free recall of responses was greater for those responses associated with high- rather than low-M stimuli, F (1, 180) = 11.21, $p < .005$. Similarly, there was a small but reliably superior free recall of representative over nonrepresentative responses even though these response terms did not differ in frequency, F (1, 180) = 3.89, $p < .05$. These latter results are of secondary interest since the principal importance of the free-recall data is their retention loss over delay intervals given the temporal invariance of recognition.

Aided recall. A plot of the aided-recall data is presented in Figure 2. Regardless of stimulus M, responses which are representative of the stimuli are readily recalled with virtually no forgetting of these responses over the retention intervals, whereas forgetting of the nonrepresentative responses occurs. The superiority of recall of representative responses is quite reliable, F (1, 180) = 266.17, $p < .001$, and the Delay × Response Representativeness interaction is reliable, F (4, 180) = 3.49, $p < .01$, which reflects the difference in forgetting rates for the two types of responses.

The comparison of greatest interest here is the pattern of aided recall of responses in relation to stimulus recognition over time. Stimulus-recognition performance remains essentially invariant over time despite differences in immediate recognition for the different experimental conditions. In contrast, aided recall declines depending upon response representativeness as well as stimulus M. Taken together, these findings argue against any view which contends that stimulus recognition *at the time of recognition testing* is somehow dependent upon retrieval of the associated response terms because response recall is progressively weaker under conditions in which no loss in stimulus recognition occurs.

Recognition–Recall contingencies. Since response recall is weakened under conditions in which stimulus recognition remains stable, the evidence is against a simple mediated account of stimulus recognition. To augment this interpretation, conditional probabilities involving recall and recognition were examined. Figure 3 presents the conditional probabilities and reveals two important features. First, there is a systematic and reliable decline in the proportion of correct response recalls given correct recognition of the corresponding stimuli, F (4, 180) = 44.88, $p < .001$. This decrease occurs for both the free- and aided-recall measures although the decrease is faster for free than aided recall as reflected in the reliable Task (free vs. aided recall) × Delay interaction, F (4, 180) = 7.90, $p < .001$. This simply indicates that the relative importance of the stimulus item as a retrieval cue increases with the length of the retention interval. More importantly, the steady decrease in these conditional probabilities, especially aided response recall given stimulus recognition, argues against the notion that stimulus recognition is in some way mediated by recall of its response mate at the time of recognition testing.

The second interesting finding in Figure 3 concerns the reverse contingency, the proportion of correct stimulus recognitions given either free or aided correct recalls of the responses. A nonsignificant delay effect, F (4, 180) = .57, $p > .05$, indicates no systematic change in this measure across retention intervals. The probability of correctly recognizing a stimulus given recall of the response was .71 at the immediate test condition and .72 after 28 days, and while the contingency is moderately high, it reveals simply that neither free nor aided recall of a response guarantees perfect correct recognition of its stimulus mate.

DISCUSSION

Three findings are of interest in this study. The first concerns the interaction between stimulus M and response representativeness on stimulus recognition. In the case of high-M trigram doublets, stimulus recognition was independent of associative training with representative responses; in contrast, with low-M stimuli, response representativeness clearly facilitated stimulus recognition. This interaction is in accord with the conceptual coding hypothesis (Ellis, 1972; Ellis & Muller, 1964) if the M value of the stimuli are regarded as one index of their codability. High-M stimuli are readily encoded and hence do not benefit from any unique encodings suggested by the representative response. On the other hand, low-M stimuli are less easily encoded, and therefore, encodings produced or suggested by a representative response enable S to more rapidly or effectively encode the stimuli than is the case with nonrepresentative responses.

This interaction parallels stimulus-recognition findings previously reported with respect to stimulus complexity (Ellis & Muller, 1964, Experiment I), stimulus M (Ellis, 1968, Experiments I & II), degree of training (Ellis & Daniel, 1971, Experiments I & II), and number of items to be stored (Santa & Ranken, 1972, Experiment II), where visual form stimuli were employed. These four studies indicated that the relative effects of associating representative responses to stimuli on the subsequent recognition of the stimuli (relative to either associating nonrepresentative responses or merely observing the stimuli) increased with high-complexity forms, increased with low-M stimuli, increased with fewer practice trials, and increased with an increasing number of items to be placed in memory, respectively. All four operations can be seen as conditions which increase the difficulty of encoding the stimuli, thereby placing increasing demand upon memory load. These relationships are all predicted by the conceptual coding hypothesis (Ellis, 1972) which proposes that easily encoded stimuli are encoded virtually free of their context, whereas contextual factors become increasingly important with a decrease in the initial codability of the stimuli.

Bower's (1970) failure to find response effects on stimulus recognition of nouns in an associative-learning task is most likely due to the high-M property of the nouns which are already easy to encode and hence do not benefit from encodings suggested by responses. Similarly, Martin's (1967) finding that stimulus recognition was unaffected by association formation is understandable in terms of two factors. The trigram stimuli employed were only moderate in stimulus M and the response terms, digits, bore no relationship to the stimulus terms.

The more general significance of the fact that properties of an overt response can affect the manner in which a nominal stimulus is encoded during associative learning is at least twofold. First, the encoded version of a low-M stimulus following A-B (representative) associative training is not the same as the encoded version of the same low-M stimulus following A-B (nonrepresentative) training, despite the nominal identity of A in both lists. Although the entire set of features that may accompany these two encoded versions of A cannot be precisely specified here, it must follow that A as a result of A-B (representative) training contains significantly more features or attributes than does A otherwise, at least more of those features which are necessary for correct stimulus recognition. Second, this finding has obvious implications for transfer studies employing the A-B, A-C paradigm. To the extent that the encoding of A is response dependent, then it follows that the encoding of A during A-C learning may differ from whatever encoding was achieved during A-B learning. Where the encoding of A is likely to be response dependent, then the onset of A-C learning may serve to disrupt the encoding of A established in A-B training. In contrast, under conditions in which the encoding of A is minimally dependent upon the response, then the onset of A-C learning (following A-B) should minimize the likelihood that a new encoding of A will develop if the already acquired encoding will effectively cue the response. Since the encoding of low-M stimuli is much more contextually dependent, it follows that such stimuli are more likely to be recoded or altered during A-C learning. Although the evidence for this is indirect (e.g., Bryk & Kausler, 1966), direct tests of recoding nominally identical stimuli have not been promising (e.g., Goggin & Martin, 1970; Williams & Underwood, 1970). Perhaps Martin's (1971) admonition is well taken in that rough tests of recoding may provide spurious negative results. Given the clear evidence for response-dependent encoding of stimuli in this study, the concept of stimulus recoding in negative transfer paradigms should be further examined.

The second finding of interest is that stimulus M facilitates stimulus recognition. This finding has been demonstrated in several studies employing shape stimuli (e.g., Ellis, 1968; Ellis et al., 1966; Feuge & Ellis, 1969; Price & Slive, 1970) and is the case with both single-stimulus and forced-choice measures as well as estimates of recognition memory obtained from signal-detection parameters. This study extends the generality of this finding to studies of forced-choice stimulus recognition of trigram stimuli. The simplest interpretation of the relationship between stimulus M and stimulus recognition emphasizes that the ease with which *S* can obtain a stable encoding or representation of a stimulus determines the accuracy of stimulus recognition. High-M stimuli are more likely to be encoded as a unit whereas the encoding of low-M stimuli is more likely to be fragmented or only partially encoded. Theoretically, the role of stimulus M can be conceptualized in terms of the number of variable encoding responses (e.g., Feuge & Ellis, 1969; Martin, 1968). Alternatively, we propose that the role of stimulus M can be viewed simply as a factor determining the rate or ease with which a stimulus may be encoded without assuming the development of variable encodings during associative training. Thus the effect of stimulus M on recognition (as well as associative learning) may be conceptualized simply in terms of stimulus processing time (or duration) required for *S* to obtain a stable encoding, with stimuli of lower M values requiring longer time to achieve such an effect. Moreover, although low-M stimuli may clearly require greater time for the development of a stable encoding, this increased encoding time *does not necessarily* imply the concomitant of increased encoding variability.

Finally, the results of the experiment provide additional evidence concerning the locus of response context effects on stimulus recognition. Although it appears unreasonable to assume that a stimulus-recognition test necessarily bypasses retrieval processes, the observed lack of a relationship between stimulus recognition and aided recall of responses argues against a mediational or retrieval interpretation of stimulus recognition *based upon recall of responses* at the time of stimulus-recognition testing. The fact that aided recall declines under test conditions in which stimulus recognition remains invariant with time, coupled with the results of the conditional probability analyses, is completely consistent with the findings reported by Ellis and Daniel (1971) with shape stimuli and extends the generality of these findings to verbal materials. These results lend support to the view that the effect of a response on stimulus recognition is during the initial encoding of the stimulus and that such effects persist during the retention intervals employed.

REFERENCES

ARCHER, E. J. A re-evaluation of the meaningfulness of all possible CVC trigrams. *Psychological Monographs*, 1960, **74** (10, Whole No. 497).

BOWER, G. H. Imagery as a relational organizer in associative learning. *Journal of Verbal Learning and Verbal Behavior*, 1970, **9**, 529–533.

BRYK, J. A., & KAUSLER, D. H. Stimulus meaningfulness and unlearning in the A-B, A-C transfer paradigm. *Journal of Experimental Psychology*, 1966, **71**, 917–920.

Daniel, T. C., & Ellis, H. C. Stimulus codability and long-term recognition memory for visual form. *Journal of Experimental Psychology*, 1972, **93**, 83–89.

Ellis, H. C. Transfer of stimulus predifferentiation to shape recognition and identification learning: Role of properties of verbal labels. *Journal of Experimental Psychology*, 1968, **78**, 401–409.

Ellis, H. C. Verbal processes in the encoding of visual pattern information: An approach to language, perception, and memory. In M. E. Meyer (Ed.), *Symposium on cognitive learning.* Bellingham: Western Washington State College Press, 1972.

Ellis, H. C., & Daniel, T. C. Verbal processes in long-term stimulus recognition memory. *Journal of Experimental Psychology*, 1971, **90**, 18–26.

Ellis, H. C., & Muller, D. G. Transfer in perceptual learning following stimulus predifferentiation. *Journal of Experimental Psychology*, 1964, **68**, 388–395.

Ellis, H. C., Muller, D. G., & Tosti, D. T. Stimulus meaning and complexity as factors in the transfer of stimulus predifferentiation. *Journal of Experimental Psychology*, 1966, **71**, 629–633.

Feuge, R. L., & Ellis, H. C. Generalization gradients in recognition memory of visual form: The role of stimulus meaning. *Journal of Experimental Psychology*, 1969, **79**, 288–294.

Goggin, J., & Martin, E. Forced stimulus encoding and retroactive interference. *Journal of Experimental Psychology*, 1970, **84**, 131–136.

Martin, E. Relation between stimulus recognition and paired-associate learning. *Journal of Experimental Psychology*, 1967, **74**, 500–505.

Martin, E. Stimulus meaningfulness and paired-associate transfer: An encoding-variability hypothesis. *Psychological Review*, 1968, **75**, 421–441.

Martin, E. Verbal learning theory and independent retrieval phenomena. *Psychological Review*, 1971, **78**, 314–332.

Martin, E., & Melton, A. W. Meaningfulness and trigram recognition. *Journal of Verbal Learning and Verbal Behavior*, 1970, **9**, 126–135.

Price, R. H., & Slive, A. B. Verbal processes in shape recognition. *Journal of Experimental Psychology*, 1970, **83**, 373–379.

Santa, J. L., & Ranken, H. B. Effects of verbal coding on recognition memory. *Journal of Experimental Psychology*, 1972, **93**, 268–278.

Williams, R. F., & Underwood, B. J. Encoding variability: Tests of the Martin hypothesis. *Journal of Experimental Psychology*, 1970, **86**, 317–324.

(Received June 5, 1972)

Journal of Experimental Psychology
1973, Vol. 98, No. 1, 79–84

ROLE OF INSTRUCTIONS IN ABSTRACTION OF LINGUISTIC IDEAS[1]

STUART KATZ[2]

University of Georgia

Using the Bransford and Franks paradigm, *S*s were asked (*a*) whether they had actually encountered each of a list of sentences presented earlier or (*b*) whether these sentences meant exactly the same thing as sentences encountered earlier. Results demonstrated that the linear relationship between the number of ideas in a sentence and recognition-confidence ratings was significantly reduced in the exact meaning condition. It was suggested on the basis of these findings and others that the linear effect in the Bransford and Franks studies is due primarily to a particular instructional set and is irrelevant to the study of uniquely semantic processes.

Several recent studies have demonstrated that semantic information in sentences and connected discourse is stored in memory as a unified representation (e.g., Begg & Paivio, 1969; Dooling & Lachman, 1971; Fillenbaum, 1966; Pompi & Lachman, 1967; Yuille & Paivio, 1969). In an effort to determine something more specific about the structural characteristics of these representations, Bransford and Franks (Bransford, in press; Bransford & Franks, 1971) carried out a series of experiments which revealed, among other things, (*a*) that individuals are unable to distinguish between closely related sentences which they heard during an acquisition period and those not actually heard, and (*b*) that reported recognition of these sentences is a positive linear function of their "sentence complexity," i.e., the number of simple ideas contained in them. On the basis of these findings, the authors argued that people spontaneously construct and store in memory a unified semantic representation based on a series of related linguistic inputs and that recognition of new inputs depends on the number of ideas common to the new inputs and the holistic representation.

Aside from questions of interpretation raised by these findings, one can question the way in which they were obtained. Since Bransford and Franks (1971) have been concerned to explain *semantic* processes, their interpretation of these has focused on memory for the underlying representations of linguistic inputs and not on memory for other aspects of sentences. Therefore, it seems somewhat inappropriate that during the recognition phase of their experiments they instructed *S*s to evaluate sentences only according to whether such sentences had *actually* been heard during acquisition. There is little doubt that under such circumstances *S*s are attentive not only to the meaning of sentences, but to various aspects of their form and the context in which they were presented. In particular, it seems likely that *S*s have trouble remembering whether they encountered a given simple idea alone or in one of many possible combinations, even though they are certain that it *means* the same thing as something they have heard earlier.

The purpose of the present study was to determine whether the Bransford and Franks (1971) results can be obtained when nonsemantic criteria are disregarded and recognition is based only on meaning. This was achieved by replicating the Bransford and Franks study using two different sets of instructions. In one condition, *S*s were instructed to rate sentences according to whether they had actually been encountered during acquisition. This Actual Instructions (AI) condition may be viewed as a straightforward replication of the original research. In the Meaning Instructions (MI) condition, *S*s were told to rate each sentence according to whether it *meant exactly the same thing* as some sentence heard earlier.

---

[1] This article is based on part of a PhD thesis submitted to the University of Michigan in 1971. The research was supported by a Ford Foundation Predoctoral Fellowship and by National Institute of Mental Health Grant MH-16950-03. The author thanks Eugene Burnstein for his invaluable advice and support, and Judith Reitman, Dorwin Cartwright, Robert Bjork, and Daniel Green for helpful criticisms. The author also thanks Phillip Shaver for his criticisms of the present article and Ernst von Glasersfeld for his valuable suggestions.

[2] Requests for reprints should be sent to Stuart Katz, Department of Psychology, University of Georgia, Athens, Georgia 30601.

If the Bransford and Franks (1971) semantic interpretation is correct, results similar to theirs should be obtained under both conditions. Specifically, (*a*) *S*s should not be able to discriminate sentences heard during acquisition from novel sentences in either the AI or the MI conditions; (*b*) a positive linear relationship between sentence complexity and recognition should occur, and the slopes of the curves for the two conditions should not be significantly different.

As a check on the success of the instruction manipulation, two special classes of test sentences were added to the recognition list. The first, called "passives," were passive grammatical forms of four-idea sentences (based on shorter sentences actually encountered during acquisition) which were quite similar, if not identical, in meaning to their active counterparts. Since it was predicted that in the AI condition *S*s would be concerned about the structure of a recognition sentence as well as its meaning, passives should receive high negative recognition-confidence ratings because their structure is noticeably dissimilar from that of the other recognition sentences. In the MI condition, on the other hand, passives should receive high positive ratings, comparable to those received by four-idea sentences in the active voice, because the configurational aspects of passives are irrelevant to the task.

The second class of special sentences were the "noncases," employed previously by Bransford and Franks (1971). This type of sentence is incompatible with the meaning of the complete representation stored in memory, although it does not differ substantially in form. Noncases were used for the following reason. In the MI condition, *S*s might simply assume that they need not pay close attention to changes in the recognition sentences and might thus relax their criteria for assigning ratings. Should *S*s interpret meaning instructions in this way, a "positivity response bias" would be expected to affect all recognition sentences. Since it has already been shown that certain sentences, in particular the most complex or four-idea sentences, receive ratings approaching the upper end of the "recognition-confidence" scale (see Results section for a description of this dependent measure), a positivity response bias would most affect the ratings of sentences with fewer ideas. The result would be artifactually and uniformly high recognition ratings. Ratings of noncases will indicate whether the positivity response bias exists; for if it does, *S*s in the MI condition will rate noncases more positively than *S*s in the AI condition, since the former will be less discriminating in detecting differences of any kind. If noncases receive comparable negative ratings in both the MI and AI conditions, however, there is little reason to believe that *S*s in the MI condition are paying less attention to changes in recognition sentences.

METHOD

Subjects

The *S*s were 26 volunteers from two undergraduate psychology courses at the University of Georgia.

Materials

Stimulus materials consisted of a set of four complex embedded English sentences, each containing four unambiguous ideas. Three of the four complex sentences were similar to those used by Bransford and Franks (1971). They are as follows: (*a*) *The rock which rolled down the mountain crushed the tiny hut at the edge of the sea.* (*b*) *The barking dog chased the brown cat which jumped on the girl.* (*c*) *The ants in the kitchen ate the sweet jelly which was on the newspaper.* (*d*) *The thunder crashing through the valley shook the tinkling bells in the corner.* Each complex sentence was analyzed into four simple declarative sentences. These latter sentences expressed, in appropriate grammatical form, a single semantic idea. Thus, *The rock which rolled down the mountain crushed the tiny hut at the edge of the sea* was partitioned into *The rock rolled down the* *mountain*, *The rock crushed the hut*, *The hut was tiny*, and *The hut was at the edge of the sea*. These simple declarative sentences (ONEs) were then recombined into new sentences expressing two ideas (TWOs), three ideas (THREEs), or four ideas (FOURs). The FOURs, of course, were the original sentences. Three THREEs, four TWOs, and four ONEs were generated from each original FOUR in this manner, and these were used to construct acquisition and recognition lists. These 12 sentences will be referred to as an "idea group." The acquisition list contained two ONEs, two TWOs, and two THREEs selected randomly from their respective pools of ONEs, TWOs, and THREEs in each idea group. This resulted in a list of 24 sentences. With the constraint that no 2 sentences from the same idea group should appear consecutively, the 24 sentences were randomized for presentation during acquisition.

The recognition list was constructed as follows: Of the six remaining sentences in each idea group not selected for acquisition, a single ONE, TWO, THREE, and FOUR were picked for recognition ("new" sentences). A single ONE, TWO, and THREE belonging to the same idea group were added from the acquisition list ("old" sentences). Thus, 28 sentences were selected for the recognition list. As in the selection procedure for acquisition, when there was more than one to choose from, selection was carried out randomly. To this list were added two passives and two noncases. A passive was defined as a FOUR which had undergone a passive grammatical transformation. The transformation changed the structure of the FOUR considerably, but did not substantially change its meaning. The passives were: (*a*) *The tiny hut at the edge of the sea was crushed by the rock which rolled down the mountain*. (*b*) *The brown cat which jumped on the girl was chased by the barking dog*.

A noncase was defined as a FOUR in which relationships among grammatical units were changed so as to change the meaning of the sentence. This was done by switching dependent clauses or prepositional phrases within the original FOURs. The noncases were: (*a*) *The barking dog which chased the brown cat jumped on the girl*. (*b*) *The ants on the newspaper ate the sweet jelly which was in the kitchen*.

With the addition of two passives and two noncases, the recognition list contained, in all, 32 sentences. Again with the constraint that no two sentences from the same idea group appear consecutively, the sentences were randomly ordered for presentation to *S*s during recognition.

Procedure

The procedure closely paralleled that used by Bransford and Franks (1971). The experimental session was carried out in two phases, acquisition and recognition.

Acquisition. Acquisition involved an incidental learning task. The *S*s, in groups of two or three, first listened to sentences read by *E* and followed by a 4–5 sec. interval in which *S*s counted out loud the number of fingers on *E*'s right hand. After this interpolated task, *E* asked a very brief question about the sentence and *S*s wrote down their answers. This procedure was repeated for each of the 28 sentences on the acquisition list. Appropriate counterbalancing was used to control for possible order effects in list presentation. When the list was completed, *S*s were allowed to rest silently for 5 min. During acquisition no mention was made of any recognition test.

Recognition. During recognition *S*s were told that a new set of sentences would be read to them and that these sentences were closely related to the set of sentences presented in acquisition. *In the AI condition their task was to indicate which sentences they had actually heard before and which ones they had not, while in the MI condition they were to indicate whether or not a particular sentence had exactly the same meaning as a particular sentence heard before.* For each sentence, *S*s indicated whether they recognized it or not, according to the instructions they received, by first marking a yes–no scale and then a 5-point confidence scale ranging from "very low" to "very high" confidence. This procedure was repeated for each of the 32 recognition sentences and constituted Trial Block 1. After a brief pause, *E* continued through the list a second time, in the same order. The second reading constituted Trial Block 2. Again, counterbalancing was employed to control for order effects.

RESULTS

Data for Sentence Complexity

Noncases and passives were analyzed separately. In all analyses, the dependent variable was mean recognition-confidence score, determined as follows:

1. Each *S*'s responses to the yes–no and confidence scales for each sentence were converted into a single rating. A plus (+) was assigned to a yes response and a minus (−) to a no response. The + or − was multiplied by *S*'s numerical response to the confidence scale for that sentence. This resulted in an 11-point recognition-confidence scale ranging from −5 to +5.

2. Each of these ratings was then pooled for (*a*) FOURs, THREEs, TWOs, and ONEs, (*b*) noncases, and (*c*) passives. Inspection of the data revealed no differences, in either the AI or MI condition, between recognition sentences that appeared in acquisition and those that did not (old vs. new sentences). Therefore, the old–new factor was collapsed for subsequent analyses.

Sentence Complexity

Results for sentence complexity appear in Figure 1. The data were cast into a mixed three-factor design with instructions (MI vs. AI), trial blocks (1 and 2), and sentence complexity (FOURs, THREEs, TWOs, and ONEs) as the factors. The overall analysis of variance yielded significant F ratios for sentence complexity, the Instructions × Sentence Complexity interaction, and the Instructions × Trial Blocks × Sentence Complexity interaction. A linear trend analysis was performed to determine whether the slopes of the sentence-complexity curves under AI and MI conditions were different. Results were as follows:

1. A significant linear trend was found for sentence complexity, F (1, 24) = 61.73, $p < .001$. It can be seen in Figure 1 that the main contribution to this effect comes from the AI condition, which clearly replicates the findings of Bransford and Franks (1971).
2. A significant interaction was found between instructions and the linear trend for sentence complexity, F (1, 24) = 36.30, $p < .001$. As Figure 1 shows, the curve for sentence complexity in the MI condition was considerably higher than the curve in the AI condition and its slope was virtually zero.
3. The linear second-order interaction of Instructions × Trial Blocks × Sentence Complexity was significant, F (1, 24) = 5.56, $p < .05$. Thus, the difference between the slopes of the MI and AI curves became more pronounced over trial blocks. Inspection of Figure 1 shows that the MI curve was even flatter and the AI curve steeper on Trial Block 2.

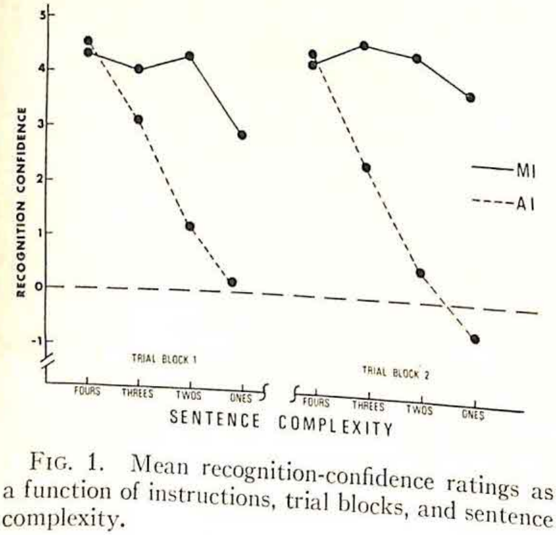


FIG. 1. Mean recognition-confidence ratings as a function of instructions, trial blocks, and sentence complexity.

In addition to the effects revealed by statistical analysis, it is also readily apparent from Figure 1 that ratings for FOURs are as high or higher than ratings at all other levels of sentence complexity in both MI and AI conditions. The finding confirms that *S*s integrate related ideas into a unified semantic representation.

Passives and Noncases

Data for each of the two special types of sentences were analyzed using a mixed design with instructions and trial blocks as factors. Results for passives revealed a significant main effect for instructions, F (1, 24) = 96.92, $p < .001$. Mean ratings, collapsed over trial blocks, were 4.48 and −2.96 in the MI and AI conditions, respectively. Thus, *S*s in the MI condition were very confident that passives meant the same thing as some sentences presented during acquisition, while *S*s in the AI condition were quite confident that they had not heard passives earlier. Results for noncases showed no significant effects. Mean ratings, collapsed over trial blocks, were −2.13 and −1.17 in the MI and AI conditions, respectively. Therefore, MI *S*s were as confident that they had not seen noncases as were AI *S*s.

Incidentally, it should be noted that when the data for the experiment as a whole were reanalyzed using percentage of yes responses instead of recognition-confidence ratings, the pattern of results and statistical analyses proved identical. This indicates that the confidence rating is a gratuitous measure.

DISCUSSION

Before the main findings of the study can be discussed in terms of instructional differences, it is necessary to know whether the instructions were properly understood during the recognition phase of the experiment. It will be recalled that passives and noncases were included in the recognition list for this purpose. Consider first the passives. It was expected that in the AI condition *S*s would attend to meaning and to structural aspects of the sentences, while in the MI condition they would pay attention only to meaning. The recognition-confidence ratings for passives support that prediction. In the AI condition, passives received the lowest ratings of any of the sentences. Since passives differ considerably in structure but very little in meaning from other sentences, *S*s must have been attending to structure as well as meaning. In the MI condition, on the other hand, passives received very high positive ratings, showing that structural differences played little or no role in recognition. Thus, the pattern of results for sentence complexity obtained in the MI condition cannot be attributed to differences in sentence form.

It still might be argued, however, that the MI condition merely gave rise to a positivity response bias, implying that *S*s were paying less attention to any and all distinguishable differences between sentences. If this were true, noncases, which differ grossly in meaning but little in form from other sentences, should be detected less frequently by *S*s in the MI than in the AI condition. In other words, recognition-confidence ratings for noncases in the MI condition should be significantly more positive than those in the AI condition. The results for noncases rule out the existence of such a bias. Indeed, the observed trend, although not statistically significant, indicates that *S*s were better at detecting noncases in the MI condition.

Effect of Instructions

It was suggested earlier that the Bransford and Franks (1971) findings were based on a particular type of instructions and consequently that their results may have little to do with meaning. The purpose of the present study was to determine whether, under meaning-orienting instructions, their findings could be replicated. The results show that this is only partly the case. On the one hand, in both AI and MI conditions, FOURs received ratings as high or higher than all other levels of sentence complexity; furthermore, *S*s had difficulty discriminating old from new sentences. However, the relationship between sentence complexity and recognition-confidence ratings changed considerably in the MI condition. The sentence-complexity curve was essentially flat in this condition; the ratings for FOURs, THREEs, and TWOs were practically identical. While the ratings for ONEs appear to be somewhat lower than the ratings for other levels of sentence complexity in the MI condition, an a posteriori comparison of the ONEs against the average of the other levels in that condition (collapsed over trial blocks) revealed no significant differences ($p > .20$). The important comparison between the MI ONEs and the AI ONEs was highly significant ($p < .001$).

Furthermore, the finding that the MI curve becomes flatter and the AI curve steeper over trial blocks is additional evidence that reliable and fundamentally different processes operate under the two conditions. Under the circumstances, it appears that further information made available on Trial Block 1 enhances different processes already initiated during acquisition.

How are we to interpret these findings? Obviously, the results show that the linear effect is not replicable across all types of instructions relevant to the task. This is interesting in its own right but not very revealing. A possible explanation might be as follows: Under MI instructions, *S*s are able to generate semantic representations from the holistic idea to match the representations of inputs, resulting in high ratings across all levels of sentence complexity. Or, and this may be another way of saying the same thing, *S*s can determine whether the representation of an input matches any part of the complete idea; this would again result in high ratings across all levels of sentence complexity. In either case, the outcome confirms what common sense would tell us: If an individual has fully integrated and memorized a complex idea, he can "recognize" any part of it. In light of this, the explanation for the linear effect offered by Bransford and Franks (1971), namely that recognition is a function of the number of basic ideas in the integrated idea exhausted by the semantic representation of the input, is unrealistic. If true, it would mean that individuals are literally unable to recognize small components (basic ideas) of a holistic idea in isolation from that whole. As indicated above, both common sense and the results of the MI condition indicate otherwise.

Additional empirical evidence indicates that the linear effect is not a function of semantic processes. For example, Katz (1971) obtained the effect using highly abstract sentences which *S*s found nearly incomprehensible (e.g., *The original event created an unusual state in the structure of the system.*). In another recent study based on the Bransford and Franks (1971) paradigm, Reitman and Bower (in press) used groups of letters and numbers instead of sentences and still obtained a significant linear relationship between the number of items appearing together and recognition-confidence ratings. Thus, although the exact cause of the linear effect remains unknown, it would appear to be an artifact of procedure rather than anything central to the study of semantics.

REFERENCES

BEGG, I., & PAIVIO, A. Concreteness and imagery in sentence meaning. *Journal of Verbal Learning and Verbal Behavior*, 1969, **8**, 821–827.

BRANSFORD, J. D. The abstraction of linguistic ideas: A review. *Cognition: International Journal of Psychology*, in press.

BRANSFORD, J. D., & FRANKS, J. J. The abstraction of linguistic ideas. *Cognitive Psychology*, 1971, **2**, 331–350.

DOOLING, D. J., & LACHMAN, R. Effects of comprehension on retention of prose. *Journal of Experimental Psychology*, 1971, **88**, 216–222.

FILLENBAUM, S. Memory for gist: Some relevant variables. *Language and Speech*, 1966, **9**, 217–227.

KATZ, S. The role of representational information and instructional set in the integration of semantic ideas. Unpublished doctoral dissertation, University of Michigan, 1971.

POMPI, K. F., & LACHMAN, R. Surrogate processes in the short term retention of connected discourse. *Journal of Experimental Psychology*, 1967, **75**, 143–150.

REITMAN, J. S., & BOWER, G. H. Storage and later recognition of exemplars of concepts. *Cognitive Psychology*, in press.

YUILLE, J. C., & PAIVIO, A. Abstractness and the recall of connected discourse. *Journal of Experimental Psychology*, 1969, **82**, 467–471.

(Received June 5, 1972)

Journal of Experimental Psychology
1973, Vol. 98, No. 1, 85–90

TAPPING RATE AND EXPECTANCY IN SIMPLE REACTION TIME TASKS

P. A. VROON[1] AND A. G. VROON

Psychological Laboratory, University of Utrecht, Netherlands

A simple reaction time (RT) experiment on the basis of unpredictable foreperiod durations and of a regular stimulus series was combined with a subsidiary task, i.e., tapping short subjective time units. It appeared that the tapping rate remained constant in the unpredictable series which was accompanied by long reaction times. The regular stimulus sequence showed an assembly of wavelike signal preparations, indicated by a considerably decreasing tapping rate just before stimulus presentation. Furthermore, the duration of the last tapping interval correlated negatively with RT. Hence, the expectancy process underlying reaction time seems to have a more or less direct observational basis in the form of a consciously adopted tapping strategy. It is concluded that the tapping rate in a double-task situation presents qualitative, as well as quantitative, information about the nature of expectancy as a function of the temporal structure of the stimulus series.

Most studies on simple and choice reaction time (RT) are described in terms of an expectancy process. It is well known that RT depends on the duration of the foreperiod, i.e., on the interval between two successive stimuli, or between a warning signal and the imperative stimulus. When constant foreperiods are used, RT tends to increase when the foreperiod is prolonged (Klemmer, 1956). This effect can be ascribed to the poor timing accuracy for longer intervals (Michon, 1967). Klemmer states that these intervals are accompanied by a considerable amount of *time uncertainty*. Increasing time uncertainty by using randomized foreperiods of different durations (rectangularly distributed) within one series also increases RT. However, in the latter case this increase is more pronounced for the short foreperiods, irrespective of the range of foreperiods used (Karlin, 1959). Reynolds (1966) explains this phenomenon by assuming that time uncertainty decreases gradually during the foreperiod since the flow of time carries information about the moment of stimulus arrival. Consequently, there is an increasing expectancy level and shortening of RT.

Altogether the pertinent literature suggests that RT is some function of the expectancy level and that this factor is highly related to timing accuracy and time perception. It has, however, to be emphasized that the evidence for the *nature* of expectancy is indirect since this concept has no immediate observational basis in terms of the subject's behavior. One could, for example, assume that expectancy consists of a series of more or less discrete preparations (peaks) or that the expectancy level is subject to slow fluctuations only. This point may be illustrated by referring to Näätänen (1971) who was confronted with the problem and remarked: " . . . as long as the means for measuring expectancy directly are lacking, a definite answer to this question remains to be given [p. 321]."

A combined study of expectancy and time perception is of interest from several points of view. Measuring temporal phenomena (e.g., by asking the subject to tap subjective time units) might deliver information on the moment-to-moment changes of cognitive processes such as signal preparation. Apart from this, expectancy is a relevant concept within the cognitive theory of time perception. Michon (1966), Ornstein (1969), Vroon (1970), and Burnside (1971) emphasize the importance of information processing and memory factors with respect to the apparent duration of a task. The expectancy process belongs to this approach although no research has been published on the relationships between expectancy and time experience so far. Consequently, definite hypotheses about correlations cannot easily be formulated. Correlations between cognitive processes and time perception have been reported by Michon (1966) and Vroon (1970) who showed that apparent time is an inverse function of the information transmission rate. This tendency has been observed by analyzing tapping rates, as well as on the basis of retrospective evaluation of the total task durations. It seems reasonable to assume that there is also a connection between *some* aspect of phenomenal time and cognitive processes such as expecting signals in, for example, reaction time tasks.

---

[1] Requests for reprints should be sent to P. A. Vroon, Psychologisch Laboratorium, Varkenmarkt 2, Utrecht, Netherlands.

The aim of the present study is to find an observable correlate of expectancy only. It is of importance to realize that generating time units in a RT task does not mean by definition that the *S*'s *time experience* is measured. It may be assumed that in a double-task situation consisting of tapping and reacting to stimuli which are predictable or unpredictable in respect of time, *S* is confronted with a trade-off problem and tries to adopt a particular tapping strategy. It seems reasonable to hypothesize that *S* pays attention to the tapping task during the period he definitely does *not* expect the stimulus, but that attention switches to the RT task as soon as he starts to expect the presentation of the stimulus. Fluctuations of the tapping rate during foreperiods might present information about the trade-off between both tasks and, hence, about the nature of the expectancy process. In the case of a regular stimulus series, it may be assumed that this switch of attention is rather sudden after *S* has learned to predict the moments of stimulus arrival so that the tapping rate changes in the form of a step function. More specifically, the rate will slow down since *S* is inclined to reduce the probability that his motor action (leading to a tap) coincides with the presentation of the stimulus. This effect will appear particularly when both tasks are performed with the same hand. However, if the moments of stimulus delivery are completely unpredictable, the expectancy level will remain constant if *S*'s behavior is in accordance with the mathematical structure of the time series, and the same will hold for the tapping rate since *S* cannot base himself on any effective tapping strategy. Summarizing, our hypotheses are: (*a*) *S* tries to discover the temporal structure of the stimulus series and adopts an appropriate tapping strategy, (*b*) the fluctuations of the tapping rate present information about the specific nature of the expectancy process as a function of the temporal structure of the stimulus series, (*c*) the tapping rate just before stimulus arrival reflects the expectancy level and forms a predictor of RT.

METHOD

Subjects. Fifty students enrolled in an introductory psychology course served as *Ss*. They were split up into an experimental group of 40 *Ss* and a control group of 10 *Ss*.

Procedure. Two RT conditions were employed in both of which the stimuli functioned as warning signals for the succeeding stimuli. Task A consisted of 38 clicks of 30 msec. each presented binaurally by headphones at a regular rate of one signal per 11 sec. Task B contained an equal number of stimuli separated by unpredictable interstimulus intervals (ISIs) between 5 and 27 sec. The minimum value of 5 sec. was necessary in view of the tapping task. The interval durations were calculated in accordance with the Bernouilli theorem (earlier employed by Nickerson, 1967) and randomized over the total length of the task (407 sec.). Applying the Bernouilli theorem to the construction of a time series implies that the probability of signal presentation remains constant, irrespective of the amount of time elapsed since the last stimulus. Hence, the flow of time does not carry information about the moments of stimulus presentation. The intervals were calculated as indicated by Feller (1957):

$$F_k = Np(1 - p)^{k-1}$$

in which: F_k = number of k-sec. intervals (ISIs) in the series, N = total number of stimuli (38), p = probability that a signal will be presented at each second, k = ISI in seconds. Figure 1 shows the theoretical and employed ISIs. Both series (A and B) took an equal amount of time, and they were performed by each *S* in a scheme counterbalanced as to order of presentation. A payoff matrix for RT was introduced to ensure the subjective significance of the RT task over the tapping task. The *S* was requested to tap at a rate of o

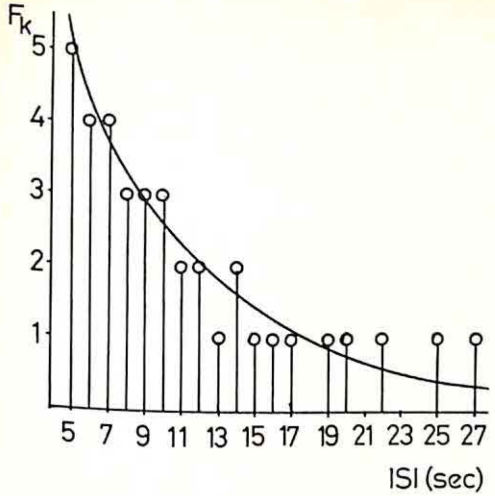


FIG. 1. Theoretical (solid line) and employed (bars) ISI frequencies.

tap per 2 subjective sec. during the tasks by pressing a microswitch mounted 1 in. next to the RT key. Both switches were operated by the forefinger of the dominant hand. The control group performed both tasks (RT on A and B and tapping) separately. The generated time units were recorded on tape and analyzed by a computer (Digital PDP 8-I).[2]

RESULTS

Reaction time. Figure 2 shows the mean RTs as functions of trials of the A and B tasks. The influence of the task structure on RT was tested by means of an analysis of variance (linear model, Tukey, 1949). Although the variances appeared to be homogeneous (Hartley [in Winer, 1970], $F = 1.20$, $F_{.05} = 1.95$), the p values were halved because of nonnormality of the distributions (goodness of fit test, $\chi^2 = 20.21$, $p \leq .01$). In spite of this, an analysis of variance may be employed since F is not very sensitive to the shape of the criterion distribution. Table 1 demonstrates that the difference between A and B is highly significant. However, there is no interaction, and the linear model appears to be valid since the nonadditive component is not significant. Table 2 presents the mean RT values of every condition and order of presentation as well as t tests on differences between pairs. It appears that

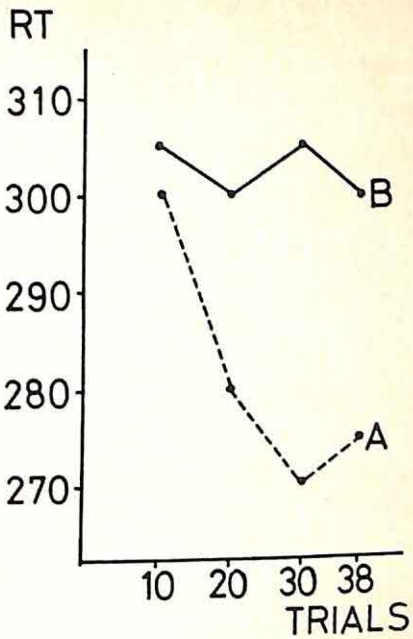


FIG. 2. Mean RTs (in msec.) over ISIs as a function of trials in the A and B tasks (experimental group, smoothed curves).

the mean RTs of B are longer than in the A series, whereas both exceed $B_{control}$ and $A_{control}$. These data indicate that the double task increased RT (performance impairment), but there are no effects of presentation order. Preparation during the B series was less effective than in the A task, but the RT values in themselves cannot clarify the (expectancy) differences which are responsible for this effect. The regression equations of four RT series as a function of trials are represented in Table 3. The B function is stationary, whereas under the A condition, RT decreases significantly with trials. The control group shows similar data so that

TABLE 1

INFLUENCE OF TASK STRUCTURE ON RT

Source	*df*	*MS*	*F*
Subjects	39	6713	
Tasks (A, B)	1	7296	13.19*
S × T	39	553	
Nonadditivity	1	715	1.30
Balance	38	548	

* $p \leq .025$.

---

[2] The authors wish to thank H. van Doorne of the Institute for Perception TNO, Soesterberg, for his intensive help in the computer analysis of the data.

TABLE 2

MEAN RT VALUES OF ALL SERIES AND t TESTS ON DIFFERENCES BETWEEN PAIRS

Series	$\overline{RT}$	Series	$\overline{RT}$	t
A	285.4	B	304.5	−3.63*
A(AB)	283.4	A(BA)	313.4	−1.65
B(AB)	295.7	B(BA)	287.4	.43
A	285.4	$A_{control}$	216.8	4.69*
B	304.5	$B_{control}$	200.0	8.26*
$A_{control}$	216.8	$B_{control}$	200.0	1.20

* $p \leq .01$.

the discrepancy between A and B cannot be ascribed to the fact that the experimental group performed a double task. The decrease of RT during the A series seems reasonable when taking into account that in the B task the ISIs remained unpredictable, whereas during A the subject had the possibility to learn to predict the moments of stimulus arrival within certain limits.

Earlier it was argued that in the unpredictable sequence RT should not change as a function of ISI duration since the flow of time does not carry information about the moments of stimulus arrival (Figure 3). A regression analysis (Hays, 1963) reveals that the curve is stationary, and these findings are in agreement with data reported by Näätänen (1971). Apparently, Reynolds' law (1966) particularly holds for a limited number of different ISIs. One could assume that this difference finds its origin in the fact that S learns to formulate *hypotheses* about the intervals between the stimuli if two or three rectangularly distributed foreperiods of considerably different durations are employed. However, such a strategy is impossible in the case of a Bernouilli sequence.

TABLE 3

REGRESSION EQUATIONS OF FOUR SERIES AND t TESTS FOR m

Series	r_{xy}	m	k	t
A	−.61	−.84	297	−4.45**
B	.18	.33	294	1.07
$A_{control}$	−.34	−.54	224	−2.11*
$B_{control}$	.17	.38	190	1.02

* $p \leq .05$.
** $p \leq .01$.

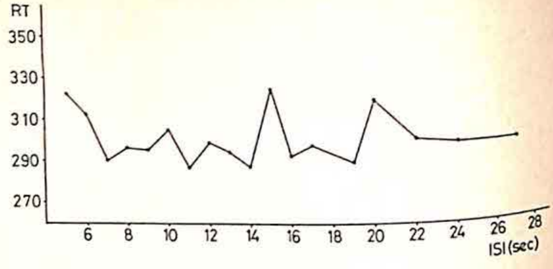


FIG. 3. Mean RTs (in msec.) in the B series as a function of ISI duration.

Tapping rate: The expectancy process. The generated time units offer a possibility to study the S's behavior continuously during the tasks. Figure 4 shows the mean lengths (over Ss) of the tapping intervals during some ISIs. It should be noted that in the case of serial production of a short interval a "lengthening effect" may occur which implies that the physical equivalent of the produced intervals tends to increase without the S's awareness (see, e.g., Vroon & Van Boxtel, 1972). The effect levels off after 40–60 productions, but it can hardly be eliminated by training (Vroon, 1972). Since this phenomenon *might* have played a contaminating role, data from the beginning of a series should not be analyzed. Apart from this, it is obvious that data after training present most information about expectancy so that the last ISIs of the A and B series were chosen (11 and 27 sec., respectively). A graph of the last 15-sec. period of the control group has been added to verify that the eventual lengthening effect had disappeared indeed.

DISCUSSION

In the A series the trapping rate is susceptible to a strong decrease a few seconds before stimulus presentation. The curve of the B series is stationary; the regression analysis (test for zero linear regression) was adopted from Hays (1963) and carried out over the raw data (40 Ss × approximately 12 taps each $= \pm 500$ measurements). In the control group (C) the function is horizontal, indicating that there was no lengthening effect at the end of the series.

It is observable that in the course of the A series, S switched his attention rather suddenly

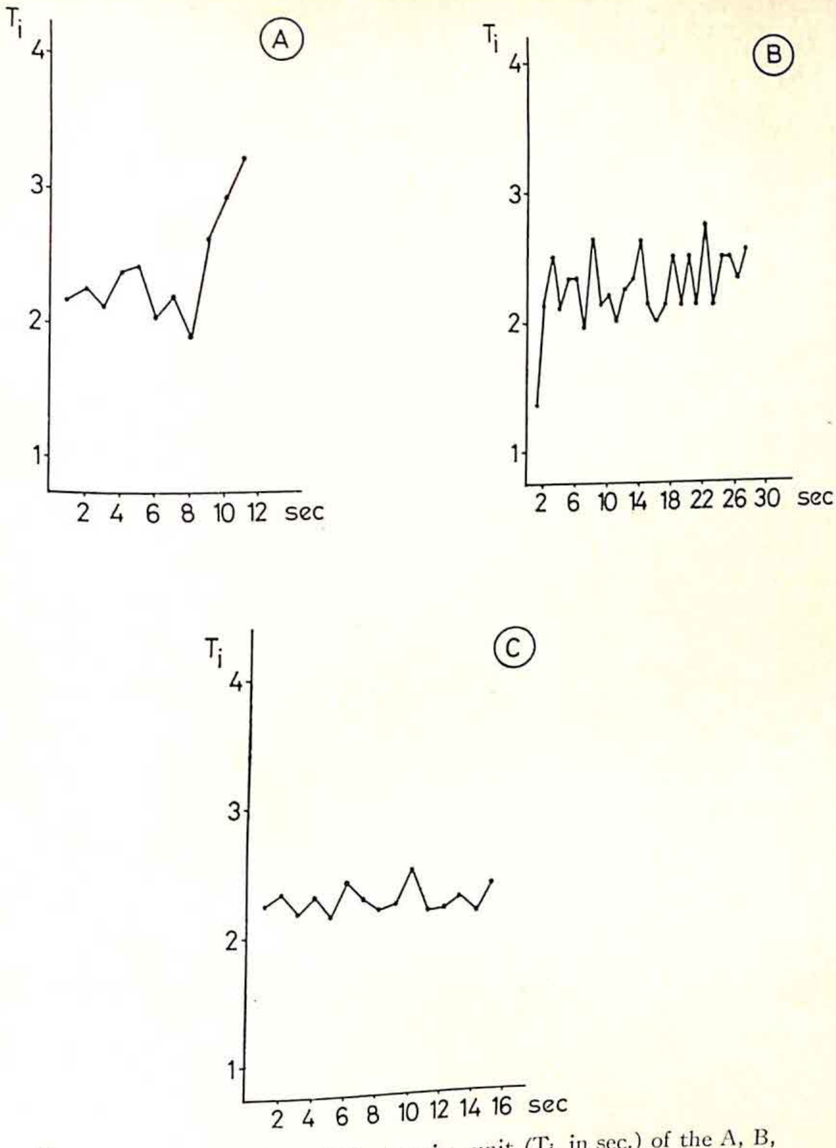


FIG. 4. Mean durations of the tapping unit (T_i, in sec.) of the A, B, and Control series.

from tapping to the RT task, i.e., he prepared himself "discretely" for the presentation of the stimulus due to the fact that he had learned to predict the moment of its arrival (see also Figure 2 and Table 3). In regular series the expectancy process seems to consist of wavelike preparations with a frequency about equal to the ISI length. The sudden change and the slowing-down effect are in agreement with the hypotheses stated in the introduction.

The B function corresponds with the curve RT–ISI length (Figure 3). The moments of stimulus arrival were unpredictable. Hence, RT should remain constant since the expectancy level cannot change systematically, and the same holds for the tapping rate. Both predictions are confirmed by the data.

However, two specific phenomena should be noted. The first RT value of Figure 3 (5 sec. ISI) is relatively high. This suggests that expectancy decreases for a short period of time after *every* reaction which is accompanied by a high RT value, and Figure 4B demonstrates that the tapping rate is very fast at the beginning of the ISI. Comparing the A and B tasks reveals that in A the tapping interval just before stimulus presentation is much longer than in B and that the mean RT of A is significantly shorter. These findings suggest that an inverse relationship between the length of the tapping interval and RT might exist. For the A series a Spearman rank-order correlation was calculated over the lengths of the last tapping interval and over the corresponding RTs of each *S* ($n = 40$). It happened that r_s reached a value of $-.266$ which is significant at the 5% level. For the B series this correlation is statistically zero ($+.03$). Thus, tapping presents qualitative information about expectancy (wavelike or constant mean level) in series with different temporal structures. Moreover, in the case of a regular stimulus sequence, RT is, to a certain degree, predictable on the basis of the tapping rate just before stimulus presentation. The most probable explanation for these findings seems to be that their basis is located in the described trade-off strategy, i.e., *S* attempts to reduce the probability that a tap will coincide with a stimulus, and he succeeds when the temporal structure of the stimulus sequence can be discovered. Any strategy, however, has to fail in the case of a Bernouilli series.

Summarizing, it may be concluded that the double-task method appears to be a useful means to study expectancy processes in a relatively direct way, and it might be possible to obtain complete "expectancy waves" as a function of different training stages during RT tasks (Vroon, in press). Furthermore, the technique might also be employed in the case of more complex information processing tasks in which expectancy is assumed to play a role.

REFERENCES

BURNSIDE, W. Judgment of short time intervals while performing mathematical tasks. *Perception & Psychophysics*, 1971, **9**, 404–406.

FELLER, W. *An introduction to probability theory and its applications.* Vol. I. New York: Wiley, 1957.

HAYS, W. L. *Statistics for psychologists.* New York: Wiley, 1963.

KARLIN, L. Reaction time as a function of foreperiod duration and variability. *Journal of Experimental Psychology*, 1959, **58**, 185–191.

KLEMMER, E. T. Time uncertainty in simple reaction time. *Journal of Experimental Psychology*, 1956, **51**, 179–184.

MICHON, J. A. Studies on subjective duration. II Subjective time measurement during tasks with different information content. *Acta Psychologica*, 1966, **24**, 205–209.

MICHON, J. A. *Timing in temporal tracking.* Assen: Van Gorcum, 1967.

NÄÄTÄNEN, R. Nonaging foreperiods and simple reaction time. *Acta Psychologica*, 1971, **35**, 316–327.

NICKERSON, R. S. Expectancy, waiting time, and the psychological refractory period. *Acta Psychologica*, 1967, **27**, 23–34.

ORNSTEIN, R. E. *On the experience of time.* London: Penguin Books, 1969.

REYNOLDS, D. R. Time and event uncertainty in unisensory reaction time. *Journal of Experimental Psychology*, 1966, **71**, 287–293.

TUKEY, J. W. One degree of freedom for nonadditivity. *Biometrics*, 1949, **5**, 232–240.

VROON, P. A. Effects of presented and processed information on duration experience. *Acta Psychologica*, 1970, **34**, 115–121.

VROON, P. A. *Some psychophysical and cognitive aspects of the time sense.* Utrecht: The University Press, 1972.

VROON, P. A. Tapping rate as a measure of expectancy in terms of response and attention limitation. *Journal of Experimental Psychology*, 1973, in press.

VROON, P. A., & VAN BOXTEL, A. Testing some implications of the sensory-physiological model of the time sense. *Psychologische Forschung*, 1972, **35**, 81–92.

WINER, B. J. *Statistical principles in experimental design.* (2nd ed.) London: McGraw-Hill, 1970.

(Received June 5, 1972)

Journal of Experimental Psychology
1973, Vol. 98, No. 1, 91–94

FACILITATION, INHIBITION, AND DISTORTIONS OF THE SERIAL CURVE IN SINGLE-TRIAL FREE RECALL AS A FUNCTION OF PRIOR WITHIN-WORD ORGANIZATION[1]

ALAIN LIEURY[2]

University of René Descartes, Paris, France

Learning of 16 bisyllabic words in single-trial free recall was preceded by repetition aloud of 40 bisyllabic pseudowords. These pseudowords were constructed from permutations of all the syllables in a set of eight bisyllabic words from the list and were called same words (SWs). The eight remaining words of the list were different (DWs) from the pseudowords. Recall scores showed facilitation for SWs and inhibition for DWs relative to recall of words (CWs) in the control group. In the control group, *S*s repeated 40 three-digit numbers prior to learning. Distortions of the serial position curve, especially a suppression of the primacy effect, appeared in SW and DW conditions. These results are interpreted in terms of transfer of a within-word plan of memorization.

Many current theories of memory insist upon the role of organization in free-recall experiments. For example, in part–whole multitrial free-recall experiments (e.g., DeRosa, Doane, & Russel, 1970, Tulving, 1966), the transfer from part-list learning to whole-list learning is interpreted as the result of *S*s having used, during whole-list learning, at least a part of the plan of memorization which had been used during part-list learnining. If the plan of memorization used for part-list learning is appropriate for whole-list learning, there is positive transfer; if the plan of memorization is inappropriate, there is negative transfer. Because in these investigations items were words, i.e., well integrated units, the plan of memorization explicitly concerns the relations between words but not the words themselves (higher order units, Tulving, 1968; retrieval plan, Slamecka, 1968). The aim of this investigation is to generalize the "plan of memorization" to the words themselves. Thus, the memorization of each word is seen as the reactivation of verbal habits or, more specifically, of the appropriate within-word plan of memorization (WWPM). The underlying idea of this conception is that weak interdependencies or conflicts between associations are not the sole causes of forgetting and that conflicts between different WWPMs are important contributing factors.

It follows from these conjectures that if the WWPM of a word is partially activated before the word is presented for learning, the retention of this word will thereby be enhanced. On the other hand, if the partially activated WWPM of a word is inappropriate for the word that follows, this might retard its integration in memory and inhibit retention.

To test this hypothesis, an experiment on transfer from an incidental learning task to a single-trial free-recall task was conducted. The free-recall list was composed of 16 bisyllabic words. The incidental task for *S*s consisted of reading aloud 40 pseudowords which were presented successively on a screen. These pseudowords resulted from new syllabic arrangements among 8 of the 16 bisyllabic words of the free-recall list and consisted of permutations between the eight first syllables of each word in this eight-word set and the eight second syllables of each of these words. Eight of the 64 possible permutations (8 × 8) were the original words, and among the 56 remaining permutations, 40 permutations without meaning were selected and were called pseudowords. For example, given the three words (in French): *index* (in English, forefinger), *merlan* (a fish), *quintal* (a French weight measure), pseudowords were *inlan*, *intal*, *merdex*, *mertal*, *quindex*, *quinlan*. The eight words of the free-recall list which were used to construct the pseudowords were called same words (SWs), whereas the eight remaining words were called different words (DWs). In the first phase of the experiment, i.e., incidental learning, the 40 pseudowords were presented successively on a screen. The *S*s were told that it was a diction exercise and their task was to read aloud, not to learn, each pseudoword as soon as it appeared. After the last pseudoword was presented, the word ATTENTION was flashed on the screen and the single-trial free-recall task began. In this second phase, the eight SWs and the eight DWs were presented alternately on the screen. The *S*s were told before the first phase that all words which appeared after the signal ATTENTION were the to-be-remembered words. When the last (16th) to-be-remembered word had disappeared, *S*s recalled all words they could from the single-trial free-recall task only.

---

[1] The author is grateful to Paul Fraisse, Geneviève Oléron, and Juan Segui for their advice; thanks are also due to Elisabeth Koskas for assisting with the preparation of this experiment.

[2] Requests for reprints should be sent to Alain Lieury, Laboratoire de Psychologie Expérimentale, 28 rue Serpente, Paris 6e, France.

According to the above hypothesis, the incidental task which produces the elaboration of WWPM partially reactivates the WWPM of the SWs. It follows that retention of SWs will be facilitated. On the other hand, the WWPM elaborated during the incidental task is inappropriate for DWs and poorer recall of these DWs is predicted. In order to go beyond the nonspecific effects resulting from the shift from the repetition task to the learning task and to estimate the degree of the specific facilitation and inhibition effects resulting from the transfer from the "pseudoword" task to the "true word" task, *S*s in a control group read aloud 40 three-digit numbers (instead of 40 pseudowords) during the repetition task. The other conditions were the same as for the experimental group. In so far as serial position effects are generally considered as resulting from different memory processes (Atkinson & Shiffrin, 1968; Tulving, 1968), SW, DW, and CW scores were studied as a function of their input positions. Because of the alternation of SWs and DWs in the free-recall learning phase, *S*s use the same rigid WWPM throughout the list. Thus, facilitation of SWs and inhibition of DWs will occur for all input positions.

METHOD

Design. Two groups, control and experimental, were used, each consisting of 32 *S*s. The counterbalancing technique was as follows: From two sets (1 and 2) of eight bisyllabic words, two sets of 40 pseudowords were constructed. Set 1 was used as SWs for half of the experimental *S*s and Set 2 was used as SWs for the remaining *S*s. In each of these subgroups and for half of the *S*s, the eight SWs were presented at each of the eight odd input serial positions and for the remaining half at each of the eight even positions. A 16 × 16 × 16 Latin square was applied to ensure that each of the 16 words of the list was used in each input position for a given subgroup of 16 *S*s. Sixteen random arrangements of pseudowords were prepared for the 16 *S*s of each subgroup. The same design was applied to the control group although the 16 words of the lists were not SWs and DWs relative to the three-digit numbers of the repetition task.

Materials. Sixteen bisyllabic words, each beginning with a different letter, were selected from a set of low-frequency words (fewer than 20/300,000; Gougenheim, Rivenc, Michea, & Sauvageot, 1964). These words had minimal phonetic and meaning similarity. Eight words were referred to as Set 1: *auteur*, *trepan*, *gamin*, *juron*, *lutin*, *niveau*, *poulain*, *radar*; and the others as Set 2: *bougon*, *engin*, *hibou*, *index*, *merlan*, *delta*, *quintal*, *sebum*.

Forty three-digit numbers were constructed from random tables. Items, words, or numbers were presented on slides. The word ATTENTION, the only word in capital letters, was presented at the same rate as the other items.

Procedure. The temporal procedure was the same for all 64 *S*s. Each *S* saw 57 slides on the screen (40 pseudowords plus ATTENTION plus the 16 to-be-remembered words); each slide was presented for 1 sec. by a Kodak Carousel projector and the time between successive slides was about 1.5 sec. The test of recall was not subject to any time restrictions and *S*s stopped when they were sure that they had recalled all items.

RESULTS

Since SWs and DWs were assigned to odd input positions for half of the *S*s of the experimental group and to even input positions for the remaining half, replotting was used to obtain complete serial position scores both for the SWs and the DWs. This operation was not used for the control group since the 16 to-be-remembered words were not SWs or DWs. In the present analysis, differences in recall scores between SW and DW conditions will be referred to as differentiation. Better recall and worse recall, relative to the control condition, will be called facilitation and inhibition, respectively. Recall differentiation between SWs and DWs was highly marked, $F\ (1,\ 93) = 149.29,\ p < .0005$; the mean SW recall score was 7.75 (maximum recall in the list = 16 words) while the mean DW recall score was 4.50. The mean CW recall score was 7.00, and the facilitation effect (SWs vs. CWs) was reliable, $F\ (1,\ 93) = 7.95,\ p < .01$; and the inhibition effect (DWs vs. CWs) was highly significant, $F\ (1,\ 93) = 88.34,\ p < .0005$. Scores for four adjacent input positions were replotted (Figure 1) in order to permit a statistical analysis of these effects for the principal parts of the serial curve.

For all parts of the serial curve, differentiation was highly marked: The three comparisons of SWs with DWs at Positions 1–8, 5–12, 9–16 were significant, $F\ (1,\ 279) = 16.66,\ p < .0005$. In the main, SWs were facilitated relative to the CWs. This effect was reliable for Positions 5–12 ($F = 3.51,\ p = .05$) but weak for the end of the curve ($F = 2.15,\ p = .10$) for Positions 9–16. On the other hand, there was no facilitation effect for the first SW relative to the first CW (the proportions of correct recall were, respectively, .438 and .688). The SWs for the four first input positions (1–4) were not better recalled than SWs for the next four positions (5–8). This shows that primacy effect for the serial curve of the SWs was suppressed.

Inhibition of the DWs relative to the CWs was more substantial. Comparison between the DWs and the CWs for Positions 1–8 yielded $F\ (1,\ 279) = 14.06,\ p < .0005$. As there was no significant difference between Positions 1–4 and Positions 5–8 for the DWs, there was no primacy effect for the DW curve. This suppression of the primacy effect was not an artifact due to the alternation of the SWs and the DWs or to the replotting procedure; the proportions of recall of the first item, which was not contaminated by the heterogeneous nature of the to-be-remembered words in the experimental group, were .250 in the DW condition and .688 in the CW condition. For Positions 5–12, comparison between the DWs and the CWs showed little inhibition effect ($F = 3.51,\ p = .05$) while it again showed a marked inhibition effect for Positions 9–16 ($F = 7.68,\ p = .005$). This marked inhibition effect was due to the very poor scores of the DWs at Positions 13, 14, and 15. (Proportions of recall for the four last DWs were, respectively, .188, .250, .375, and .875; whereas scores for the last CWs were .469, .563, .686, and .781.) These results show that recency effect for the DWs was reduced to the terminal item only.

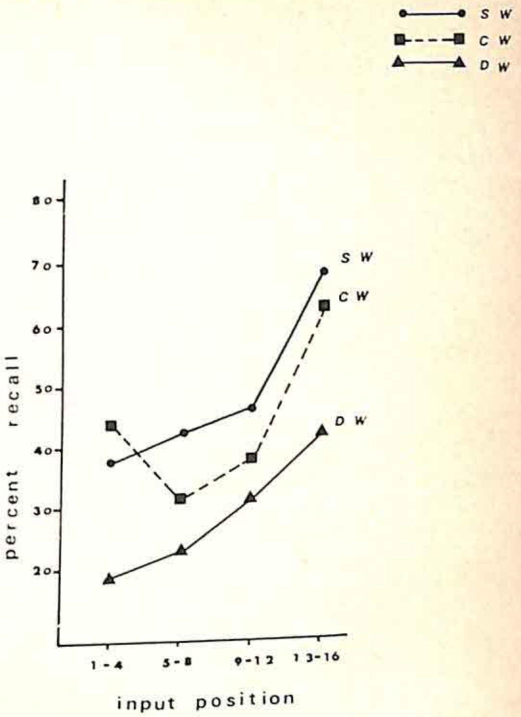


FIG. 1. Percentage of items correctly recalled in each condition as a function of serial input position.

DISCUSSION

The first hypothesis postulates the facilitation of SWs and the inhibition of DWs relative to the control condition (CWs) and the second hypothesis the generality of these effects for all input positions. The results confirm the first hypothesis, which would suggest that retention must not be interpreted only in terms of interdependencies between well integrated units. In fact, many paradigms such as paired-associate learning or multitrial learning have focused attention on the interdependencies between words (or other well integrated units such as digits or letters) which were analyzed in terms of associations and subjective units, respectively. But such interdependencies are probably weaker if they exist (Slamecka, 1968) in other paradigms, such as single-trial free-recall and short-term-memory paradigms. In light of this experiment, the specific internal organization of the words themselves must be considered as a crucial component in these cases.

The second hypothesis was confirmed in the main, except for SWs in as much as the primacy effect was suppressed; this result probably reflected a rearrangement of the pseudoword WWPMs to the SW WWPMs. Finally, the facilitation and inhibition effects also occurred for recent items. These results show that if the recency effect reflects a different memory process (short-term storage, Atkinson & Shiffrin, 1968; or a different retrieval process, Tulving, 1968), this process is not independent of organizational factors.

To conclude, what is the functional significance of the WWPM? Is it the activation of a storage organization, an encoding-retrieval plan, a phonemic program, or something entirely different? An answer to this question must await further investigation.

REFERENCES

Atkinson, R. C., & Shiffrin, R. M. Human memory: A proposed system and its control processes. In K. W. Spence & J. T. Spence (Eds.), *The psychology of learning and motivation.* Vol. 2. New York: Academic Press, 1968.

DeRosa, D. W., Doane, D. S., & Russel, B. The influence of first-list organization upon second-list free recall learning. *Journal of Verbal Learning and Verbal Behavior*, 1970, 9, 269–273.

Gougenheim, G., Rivenc, P., Michea, R., & Sauvageot, A. *L'élaboration du Français fondamental.* (Rev. ed.) Paris: Didier, 1964.

Slamecka, N. J. An examination of trace storage in free recall. *Journal of Experimental Psychology*, 1968, 76, 504–513.

Tulving, E. Subjective organization and effects of repetition in multitrial free recall learning. *Journal of Verbal Learning and Verbal Behavior*, 1966, 5, 193–197.

Tulving, E. Theorical issues in free recall. In T. R. Dixon & D. L. Horton (Eds.), *Verbal behavior and general behavior theory.* Englewood Cliffs, N. J.: Prentice-Hall, 1968.

(Received June 7, 1972)

Journal of Experimental Psychology
1973, Vol. 98, No. 1, 95–101

SCANNING FOR INFORMATION IN LONG- AND SHORT-TERM MEMORY[1]

KEITH T. WESCOURT AND RICHARD C. ATKINSON[2]

Stanford University

The *S*s memorized a list of 30 words (LT set) before participating in a recognition memory task. Test sessions were divided into pure and mixed trial blocks. During pure blocks, 1–4 words (ST set) were presented to *S* on each trial immediately before exposure of the test stimulus. The *S* made a positive response if the test stimulus was from the ST set and a negative response if it was a distractor. During mixed blocks the procedure was similar except that test stimuli were also drawn from the LT set; the *S* made a positive response if the test stimulus was from either the ST set or the LT set, and a negative response otherwise. The results indicate that in the mixed blocks reaction time to positive test stimuli from the LT set or to negative test stimuli does not depend on ST set size; for all other test stimuli, reaction time is an increasing function of ST set size. These findings are shown to be consistent with the view that *S*s can search a long-term store and a short-term store simultaneously.

Theories of memory that differentiate between a short-term store (STS) and a long-term store (LTS) lead to a consideration of how the two stores interact in situations requiring their simultaneous activity. For example, serial position effects in free-recall tasks have been attributed to interacting effects of STS and LTS. Primacy effects are attributed to increased STS processing of the earlier items of a list, resulting in more information about those items being stored in LTS, while recency effects are thought to reflect near perfect retrieval of items from the end of the list that still reside in STS at the time of test (Atkinson & Shiffrin, 1971). Such an analysis illustrates the viability of the two-store memory model, but does not specify details about the stages of processing that might be present in complex memory tasks.

Sternberg (1969) has shown that in a simple recognition task the reaction time (RT) to decide whether or not an item is a member of a small set of items in memory (six or fewer) is an increasing linear function of the number of items in the set. Sternberg accounts for these results with a four-stage model. In this model, the size of the memory set acts to influence RT by altering the time of the stage involving memory search while not affecting the stages of stimulus encoding, decision, and response. In tasks of this type, the memory search is assumed to involve comparisons of the test stimulus with the elements of the memory set in STS.

Juola, Fischler, Wood, and Atkinson (1971) report a study similar to that of Sternberg (1969) in which the memory sets were lists of 10–26 words. Each of their *S*s learned a single list prior to testing and maintained the list as the positive memory set throughout the experiment. The results were consistent with data from studies with small memory sets; that is, as list length increased the RT of both positive and negative responses increased linearly. However, the average increase per additional item in the memory set was much less (about 5 msec. per item vs. about 40 msec. per item for Sternberg). Given the long list lengths employed by Juola et al. and the fact that their procedure included a distracting counting task performed by *S*s immediately prior to each test stimulus, it may be assumed that memory search in their experimental task involved comparison of the test stimulus with information in LTS. The differences between their RT set size functions and those found by

---

[1] This research was sponsored by National Science Foundation Grant NSF G5-443X2. The authors thank James F. Juola for his comments during the planning of this study.

[2] Requests for reprints should be sent to Richard C. Atkinson, Department of Psychology, Stanford University, Stanford, California 94305.

Sternberg may therefore reflect different search rates in STS and LTS.

The tasks of both Sternberg (1969) and Juola et al. (1971) are simple in the sense that performance involves search within only one of the memory stores. A complex task using a similar procedure may be defined in which the positive memory set consists of both (*a*) a set of items learned prior to testing and held constant over test trials (LT set) and (*b*) an additional small set of items, none of which are in the LT set, presented prior to each test stimulus and varied on each trial (ST set). The total set of items requiring a positive response would thus consist of sets of items in LTS and STS, and performance would involve search of both stores. Studies of this design have been reported (Doll, 1971; Forrin & Morin, 1969); however, the LT sets used in these studies were small (less than six items) and it is not clear that *S*s did not transfer these LT sets to STS for search during the task. Indeed, Doll concluded that this is exactly the strategy employed by *S*s in his study which examined the effects of differential reward for responses to items from the two memory sets. It remains to be shown then what interactions of processing between STS and LTS will be found when constraints insure that the LT set cannot be transferred to STS immediately prior to presentation of a test stimulus.

In the present study, *S*s learned an LT set of 30 unrelated words and then participated in test sessions employing a reaction-time recognition task. Each session was divided into mixed and pure blocks of trials. On each trial of a mixed block (*a*) an ST set of 0–4 words was presented to *S*, (*b*) *S* rehearsed these words once, and (*c*) the test stimulus was presented; *S* then made a positive response if it was either in the LT set or the ST set and a negative response otherwise. For pure blocks of trials the procedure was the same except that (*a*) the ST set varied 1–4 items (no 0 condition) and (*b*) *S* was explicitly instructed that no words from the LT set would appear as test stimuli during the trial block. Thus, positive responses were required only if the test stimulus was in the ST set. The pure blocks then were essentially a replication of Sternberg's (1969) varied-set procedure in which the positive set on each trial is maintained in STS.[3] In the mixed blocks, those trials with no words in the ST set replicated the conditions of Juola et al. (1971) in that the positive set consisted of items in LTS. The trials in the mixed block with an ST set of 1–4 words introduced a condition in which the positive set consisted of information in both STS and LTS.

Reaction time was recorded as a function of trial-block type, ST set size, and test stimulus type for each trial of the test sessions. To the extent that these data are systematically interrelated, different hypotheses about interactions between STS and LTS processing can be evaluated.

METHOD

Subjects. Eighteen female students at Stanford University served as *S*s. Each *S* participated in one training session and two experimental sessions; they received $2.00 for each session.

Stimuli. Words for LT sets, ST sets, and distractor test stimuli were drawn from the Thorndike-Lorge word list. All words were 4–8 letters in length and had two syllables. Words for the training session and a block of warm-up trials for the experimental sessions had frequency counts of less than 20 per million, while words for the experimental trial blocks had frequency counts of more than 20 per million. Distractors and words in ST sets were each used on only a single trial during the entire experiment.

Three different 30-word LT sets were created. The *S*s were divided into three equal groups and each group used one of the LT sets during all sessions. The first three and last three words on each list were never tested except during the training session and were kept the same on all three lists. The middle 24 words of the lists were randomly ordered for each *S* when presented for learning prior to the first session.

Four blocks of trials were constructed for the training session and each of the two experimental sessions. Two of the blocks were pure blocks (positive test stimuli could only be drawn from the ST set presented on each trial) and two were mixed blocks (positive test stimuli could be drawn from either the ST set or the LT set). In the experimental sessions pure blocks were 16 trials in length and mixed blocks 40 trials. In addition, each experimental session started with a pure block of 12 warm-up trials, so there were 124 trials during the entire session. For all trial blocks the proportion of positive and negative trials was .5. In the experimental sessions, each different ST set size (1–4 for pure blocks and 0–4 for mixed blocks) appeared with equal probability within each trial block. In mixed blocks, for all nonzero ST set sizes half the positive test stimuli were drawn from the ST sets and half from the LT set. When combined, the two mixed blocks of each experimental session tested each of the central 24 words of an LT set exactly once.

---

[3] Sternberg's ST sets were sampled with replacement from the set of available items (digits), whereas the present study sampled items (words) without replacement.

Each *S* in the experiment received exactly the same words as test stimuli (except positive test stimuli drawn from the LT set which varied as a function of the LT set learned by *S*s). However, trial blocks presented at each experimental session and ordering of trial-block types within sessions were systematically varied across *S*s.

Apparatus. The ST sets for each trial were tape recorded. During testing, they were presented with a manually operated recorder through a speaker system.

Each test stimulus was typed in captial letters on a white 5 × 8 inch index card with an IBM Executive Registry typewriter. The stimuli were presented by means of a tachistoscope. They appeared in the center of the visual field and subtended a horizontal visual angle of about 2°. A fixation target at the center of the field was displayed for .5 sec. prior to the onset of the test stimulus on each trial. The display field was dark between trials.

Below the tachistoscope, in front of *S* and to her right on a table, three telegraph keys were arranged along an arc. The centers of adjacent keys were separated by about 3 cm. The *S* could comfortably rest her right forefinger on the center key between trials and make a short, natural movement to the left or right to strike either of the two response keys. Half of the *S*s were randomly assigned the right key as the positive response and the left as the negative response; the assignment was reversed for the other *S*s. In her left hand the *S* held a button used to initiate display of the test stimulus on each trial.

A small printed sign was attached to the side of the tachistoscope facing *S* and served as a reminder of whether the current trial block was pure or mixed. One side of the sign instructed that during mixed blocks (A blocks to *S*) words from the LT set would appear as test stimuli; the other side instructed that during pure blocks (B blocks to *S*) words from the LT set would never be tested. The correct reminder was displayed during each block of trials and was changed by *S* between blocks on instructions from *E*.

Procedure. The *S* was contacted by phone 12–48 hr. prior to the training session and given one of the LT sets to learn. The instructions stated that the list was to be learned in the order given over the phone so that it could be written and recited perfectly.

At the start of the training session *S* was asked to write the words in the LT set in the correct serial order. After a correct written recall, *S* was asked for an oral recitation of the list. All *S*s satisfied a criterion of successive correct written and oral recalls of the LT set.

The *S* was then seated at the tachistoscope and given detailed instructions about the task. The *S* was told that *E* was investigating how rapidly people can make decisions about words they have learned. The differences between pure and mixed blocks were carefully outlined emphasizing that (*a*) both words from the LT set and from the ST set required a positive response when they appeared as test stimuli in mixed blocks, (*b*) in pure blocks LT set words would never appear as test stimuli and so positive responses should only be made to test stimuli drawn from the ST set, and (*c*) on several trials in each mixed block only the command "count" would be heard from the tape recorder and that this meant *S* was to count backward "4,000, 3,000, 2,000, 1,000" and then initiate the display which, since no ST set had been given, could only be an LT set word or a distractor.

The *S* was instructed that, except for these differences, the procedure for each trial in each block would be the same: (*a*) The *E* would say "ready" and turn on the tape recorder presenting an ST set; (*b*) the *S* would either repeat the ST set once aloud or count backward if there was no ST set; (*c*) the *S* would then press the button in her left hand; (*d*) the fixation target would appear and .5 sec. later the test stimulus would be exposed; (*e*) the test stimulus would remain visible until *S* made either a positive or negative response by pressing the appropriate key; and (*f*) *E* would record the response latency and change the stimulus card (the intertrial interval was about 10 sec.). The *S* was also told that ST set words and negative test stimuli would each appear on only one trial in the experiment. The instructions emphasized that *S* was to respond as rapidly as possible while trying to avoid making errors.

The test phase of the training session then began. The *S* was told that the first day was a practice session and that she should try to find the best strategy for performing the task. During the session, feedback was given between trials about speed and accuracy of responses.

The two experimental sessions were held on subsequent days; each session began with a written and oral recall of the LT set. The instructions were then reviewed. The warm-up block and four test blocks were presented. Only feedback concerning incorrect responses was given during experimental sessions. There was a rest period of about 3 min. between trial blocks and each session lasted approximately 1 hr.

RESULTS

For mixed trial blocks, there were 14 trial types defined by combination of ST set size (0–4) and test stimulus type (ST set, LT set, or distractor). Similarly, there

TABLE 1

RT MEANS (IN MSEC.), STANDARD DEVIATIONS, AND ERROR RATES FOR ALL TRIAL TYPES

Test stimulus type	ST set size	M	SD	Percentage of errors
Pure ST-yes	1	454	57	0.0
	2	514	64	2.2
	3	556	72	1.4
	4	554	71	2.2
				$\bar{X}$ = 1.5
Pure no	1	527	81	0.0
	2	562	79	0.0
	3	599	81	.7
	4	596	81	.7
				$\bar{X}$ = .4
Mixed LT-yes	0	623	77	2.2[a]
	1	675	76	6.6
	2	690	75	8.0
	3	705	80	6.6
	4	686	62	7.3
				$\bar{X}$ = 7.1
Mixed ST-yes	1	541	84	1.5
	2	569	83	2.9
	3	586	80	.8
	4	604	83	2.9
				$\bar{X}$ = 2.0
Mixed no	0	701	93	3.2[a]
	1	731	84	6.2
	2	713	113	2.5
	3	746	111	4.7
	4	721	124	4.0
				$\bar{X}$ = 4.3

[a] Not included in computation of mean percentage of errors.

were eight trial types defined for pure trial blocks (ST set [1–4]; test item types [ST set, distractor]). For each *S*, the mean RT for correct responses for each trial type in both experimental sessions was computed.[4] The mean of the *S* means for each trial type is presented in Table 1; also given for each trial type is the standard deviation of the *S* means and error rate. In Figure 1a the mean RTs for each type of test stimulus are shown as a function of ST set size for pure trial blocks. Figure 1b presents similar data for the mixed blocks. Lines fit by the method of least squares are shown for the ST-yes and no (distractor) functions in Figure 1a and for the ST-yes function in Figure 1b. Lines representing the average RT across ST Set Sizes 1–4 for the LT-yes and no functions in Figure 1b are also shown.

In both pure and mixed blocks negative responses were slower than positives. In mixed trial blocks LT-yes responses were slower than ST-yes responses. Several of the functions show a drop in RT from ST Set Sizes 3 to 4; the meaning of this result is not clear, though it may be due to the overt rehearsal of the ST sets just prior to test. Separate analyses of variance were carried out for the five functions shown in Figure 1 using the *S* means as scores. The effect of ST set size was significant for all analyses: pure ST-yes, $F\ (3, 51) = 39.645$, $p < .01$; pure no, $F\ (3, 51) = 25.012$, $p < .01$; mixed ST-yes, $F\ (3, 51) = 6.3444$, $p < .01$; mixed LT-yes, $F\ (4, 68) = 10.418$, $p < .01$; and mixed no, $F\ (4, 68) = 3.374$, $p < .05$. However, subsequent tests on the mixed LT-yes and mixed no functions show that in both cases ST Set Size 0 differs significantly from the mean of ST Set Sizes 1–4, $F\ (1, 68) = 37.100$, $p < .01$; $F\ (1, 68) = 6.532$, $p < .01$, whereas the differences among ST Set Sizes 1–4 are not significant, $F\ (3, 68) = 1.64$, $p > .10$; $F\ (3, 68) = 2.308$, $p > .05$. Thus, responses to the LT set test items

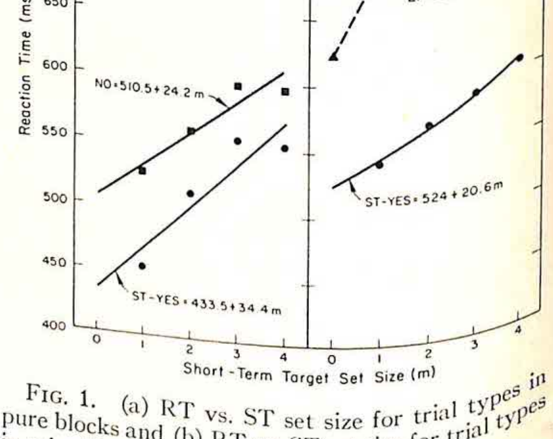


FIG. 1. (a) RT vs. ST set size for trial types in pure blocks and (b) RT vs. ST set size for trial types in mixed blocks.

---

[4] Outlying scores for each *S* were eliminated from computations by the following method: (*a*) For each trial type, the mean was calculated, (*b*) each score that was more than 1.5 times greater than the mean was deleted, and (*c*) the mean was recalculated from the remaining scores. Using this procedure, about 3.6% of the scores were eliminated.

and distractor items in the mixed trial blocks do not exhibit the systematic effect of ST set size apparent in the other functions, though there is, in these two cases, an effect of presence or absence of an ST set on a trial.[5]

The overall error rate was less than 5%. From Table 1 it can be seen that the highest error rate was a 7.1% incorrect rejection rate for LT set test items in the mixed blocks. The error data do not demonstrate any systematic effect of ST set size.

Finally, for data from the mixed trial blocks, mean RTs for each *S* for each of the three test stimulus types collapsed across ST set size were computed conditional on the test stimulus type of the previous trial. Table 2 presents means of these *S* means in a matrix form, where rows represent events on Trial $n - 1$ and columns, Trial n. A separate analysis of variance was performed for each column in Table 2; the only significant effect was found when on Trial n an LT set item was presented as the test stimulus, $F(2, 34) = 4.633$, $p < .05$. Correct responses in this case were fastest when Trial $n - 1$ also had an LT set item as a test stimulus and slowest when it had a distractor test stimulus. These results might be interpreted as evidence for a "priming" effect; retrieval from LTS somehow facilitates immediately subsequent retrieval from it.

Discussion

The present study addresses the question of how to describe search through memory when there is relevant information in both STS and LTS. We are not concerned here with distinguishing memory processes within either store (e.g., serial vs. parallel scanning of items), but only with describing how the memory stores are accessed; in particular, the question may be asked whether STS and LTS are searched simultaneously in the present task or whether they are processed sequentially with search of one store delayed until search of the other store is completed.

TABLE 2

Mean RT (in msec.) on Trial n Conditional on Trial Type of Trial $n - 1$

Trial $n - 1$	Trial n		
	LT-yes	ST-yes	No
LT-yes	665	602	762
ST-yes	685	624	738
No	707	582	739

It has been shown that various models which posit serial or parallel scanning within memory stores cannot be uniquely distinguished with the available data (Townsend, 1971). Similar comments concerning identifiability of alternative models also could be directed toward any attempt made here to formulate a single model to account for the present results. Therefore, our approach will be to consider the two simplest, most intuitive models for memory processing in the present task and to show that one of them is inconsistent with the data, while the other provides a reasonable account of the major results. We cannot, however, exclude the possibility that more complex models (e.g., those incorporating probability mixtures of alternative processes) also can be formulated to account for these data.

There are two special cases for any memory model concerned with describing tasks requiring both STS and LTS search. The first case is illustrated by the pure blocks of the present task where there is no need to search LTS in order to make a response. A representation of the information processes underlying performance in this case is shown in Figure 2a. Only STS processing occurs during the search stage; it is assumed that only the length of the search stage is affected by the number of items in STS, such that RT increases as the number of items increases. This description is comparable to but less specific than Sternberg's (1969) model for scanning STS.

The second special case is when there is no need to search STS before making a response; this occurred in the present study on those trials of the mixed blocks when no ST set was presented. A representation for this case is shown in Figure 2b. The flow chart is identical to Figure 2a except that only LTS is processed during the search stage. Again the assumption is made that only the search time is affected by LT set size. Since in the present task the

---

[5] Subsequent experiments, using a similar paradigm, have provided strong support for our interpretation of the mixed LT-yes and mixed no functions as flat across ST Set Sizes 1–4.

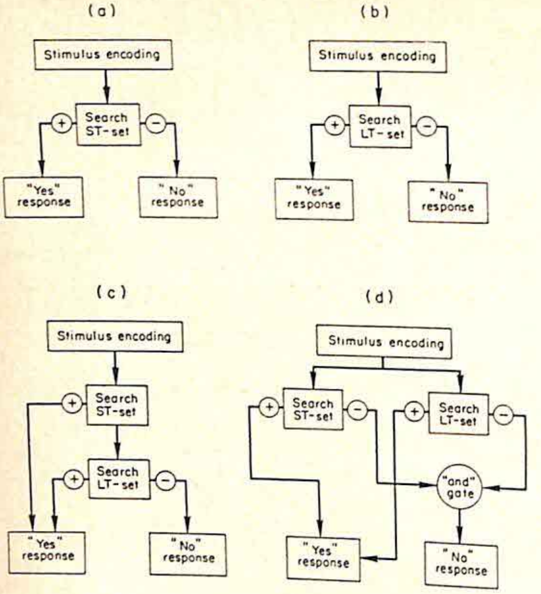


FIG. 2. Representation of processing stages when (a) memory set items are only in STS, (b) memory set items are only in LTS, (c) memory set items are in both stores and STS is accessed first, and (d) memory set items are in both stores and the stores are accessed simultaneously.

LT set size was constant at 30 words, the expected value for search time is constant.

When items to be searched reside in both STS and LTS (as in the mixed block when an ST set was presented), there are two intuitively simple models for how the two stores are accessed. One is a sequential model which postulates that only one memory store is searched at a time; the other is a simultaneous model in which search through both stores begins at the same time. In Figure 2c we represent a sequential model in which STS is always searched first. This ordering of search is consistent with the result that for all ST set sizes mean RTs for mixed ST-yes trials are faster than those for mixed LT-yes trials. In this model, if the test stimulus is located during the initial STS search, then a positive response is made immediately without searching LTS; otherwise, LTS is searched and, depending on whether or not the test item is located there, a positive or negative response is made. There are two critical predictions from the sequential model of Figure 2c for the present experiment:

1. The RT vs. ST set size functions for positive test items drawn from the ST set in mixed blocks (mixed ST-yes) and in pure blocks (pure ST-yes) should be identical. This follows because the processing stages involved in generating responses to such items in Figure 2c are isomorphic to those in Figure 2a.
2. The RT vs. ST set size functions in the mixed blocks for positive test items drawn from the LT set (mixed LT-yes) and for distractor test items (mixed no) should vary with ST set size in about the same way (i.e., have similar slopes) as the mixed ST-yes. This is the case because these two trial types involve a combined search stage that can be described as the sum of a component that varies with ST set size (STS search time) and a constant valued component (LTS search time).

In contrast, for the simultaneous model illustrated in Figure 2d, STS and LTS are entered at the same time and two independent searches are executed. Finding the test stimulus in either store results in the immediate execution of a positive response, with RT depending only on the size of the set of which the test stimulus is a member. A negative response, however, cannot be made until both STS and LTS are searched and the test stimulus not found in either store. Thus the simultaneous model incorporates the equivalent of a logical "and" gate between the search and responses stages for negative test stimuli. Predictions from this model are:

1. The mixed-no function should reflect only the longer of the search times of STS and LTS. This follows from the presence of the "and" gate which delays responses to negative search outcomes until the final search is completed.
2. The mixed LT-yes function should not reflect ST set size and should therefore have a 0 slope.

Given the above predictions, it is clear that the sequential model is inconsistent with the results of the present study. First, there are obvious, large slope and intercept differences between the pure ST-yes and mixed ST-yes functions in Figure 1. Second, the analyses of variance and subsequent comparisons on the mixed LT-yes and mixed no functions show no significant relation between RT and ST set size for these trial types, though there are significant differences in both functions for the case when ST set size was zero. These results are consistent with the simultaneous model we have outlined if the time of LTS search is greater than the time of STS search for all ST set sizes. When this is true, the simultaneous model predicts the flat lines fit to the mixed LT-yes and mixed no functions in Figure 1b. The model as presented here does not account for the slope differences between the pure ST-yes, pure no and mixed ST-yes functions or for the significant differences within the mixed LT-yes and mixed no data between zero and nonzero ST set sizes. However, a quantitative model incorporating the simultaneous model for accessing STS and LTS as described here and a two-stage LTS search process has been reported elsewhere (Atkinson & Juola, 1973, in press); this model, developed from results of several studies, can account for these specific aspects of the present results that the model presented here does not distinguish.

Returning to a study cited earlier (Forrin & Morin, 1969), further evidence can be found favoring a simultaneous model. In that study, only small LT sets (1, 2, or 3 items) were learned by *Ss*, and ST sets were of similar size range. Translating their results into the present terminology, it was found that pure ST-yes and mixed ST-yes functions varied with ST set size, but had different intercepts. The pure no and mixed no functions were similarly related. However, as in the present study, the mixed LT-yes function did not increase with ST set size. These results follow from the simultaneous model if LTS search time is faster than STS search time for all ST set sizes; given the extremely fast scanning rate for LTS vs. that for STS described by Juola et al. (1971), this case is probably satisfied in the Forrin and Morin study because of the small LT sets they employed. Thus, the simultaneous model can account for both the similarities and differences between the results of the present study and this earlier study (i.e., the flat mixed LT-yes function and increasing mixed no function).

The available data therefore lead us to conclude that in a task in which there is relevant information in both STS and LTS, access to the two stores need not be sequential. Instead, it is possible that both stores are entered simultaneously and that processing within one store is relatively independent of processing within the other.

REFERENCES

Atkinson, R. C., & Juola, J. F. Factors influencing speed and accuracy of word recognition. In S. Kornblum (Ed.), *Attention & performance.* Vol. IV. New York: Academic Press, 1973.

Atkinson, R. C., & Juola, J. F. Search and decision processes in recognition memory. In D. H. Krantz, R. C. Atkinson, R. D. Luce, & P. Suppes (Eds.), *Contemporary developments in mathematical psychology.* San Francisco: W. H. Freeman, in press.

Atkinson, R. C., & Shiffrin, R. M. The control of short-term memory. *Scientific American*, 1971, **224**, 82–90.

Doll, T. J. Motivation, reaction time, and the contents of active verbal memory. *Journal of Experimental Psychology*, 1971, **87**, 29–36.

Forrin, B., & Morin, R. E. Recognition times for items in short- and long-term memory. *Acta Psychologica*, 1969, **30**, 126–141.

Juola, J. F., Fischler, I., Wood, C. T., & Atkinson, R. C. Recognition time for information stored in long-term memory. *Perception & Psychophysics*, 1971, **10**, 8–14.

Sternberg, S. Memory scanning: Mental processes revealed by reaction-time experiments. *American Scientist*, 1969, **57**, 421–457.

Townsend, J. T. A note on the identifiability of parallel and serial processes. *Perception & Psychophysics*, 1971, **10**, 161–163.

(Received June 8, 1972)

Journal of Experimental Psychology
1973, Vol. 98, No. 1, 102–108

SAME AND DIFFERENT JUDGMENTS FOR WORD–COLOR PAIRS WITH "IRRELEVANT" WORDS OR COLORS:

EVIDENCE FOR WORD-CODE COMPARISONS[1]

FREDERICK N. DYER[2]

U.S. Army Medical Research Laboratory, Fort Knox, Kentucky

Latencies for same and different judgments resulting from comparisons between words and colors were determined with one of the pair of stimuli being both a word and a color. The irrelevant aspect of this dual stimulus bore each of five possible relationships to the stimuli that were relevant to the match. These five conditions plus two control conditions produced large differences in latencies for making comparisons. Same responses were generally faster than different responses. Correspondence between the irrelevant stimulus and the combined relevant stimulus facilitated same responses but showed no facilitation of different responses. These findings indicate that basic differences exist between same and different decision processes. In some conditions, irrelevant words delayed matches between words and colors more than irrelevant colors delayed such matches. This suggests that central comparisons between the pair of stimuli were in a form more closely related to words than to colors.

Treisman and Fearnley (1969) showed that judgments of the equivalence and nonequivalence of two stimuli took more time when each stimulus was characterized by a different stimulus attribute than when the two stimuli were characterized by the same attribute. Colors and color names were compared in a card-sorting task. An example of one of the cards for "cross-attribute" matching was the word *blue* printed in black above a series of green Xs. Since the word and color do not correspond, this particular card would be sorted into the different pile. An example of "within-attribute" matching for the attribute of color was a color name in blue printed over a series of green Xs. An example of within-attribute matching for color names was the word *red* printed twice, one above the other, with the upper word in colored ink. The last two examples would be sorted different and same, respectively. A condition with simpler stimuli for within-attribute matching (both stimuli Xs for color matches and both words in black for word matches) was not included in the Treisman and Fearnley study; but the fast times for within-attribute matching with their complex stimuli relative to cross-attribute matching with the simple stimuli described above indicated that little or no interference to within-attribute matching resulted from the competing value on the other attribute.

In addition to the shorter time for within-attribute matching relative to cross-attribute matching, Treisman and Fearnley (1969) found that cross-attribute matching was considerably delayed when the upper stimulus involved in the comparison was both a word and a color, i.e., a color name printed in colored ink. The additional delay, produced by the irrelevant aspect of these combination stimuli, occurred about equally in the condition where a color was compared to a word (CW matching) and the condition where a word was compared to a color (WC matching). The difference between the stimulus cards for CW and WC matching was in the choice of a black word or a series of colored Xs, respectively, for the lower noncombination stimulus. The irrelevant aspect of the upper combination stimulus was thus a word for CW matching and a color for WC matching.

Since anomalous situations can provide

---

[1] The author wishes to acknowledge the assistance of Laurence J. Severance who helped develop procedures and collected the data.

[2] Requests for reprints should be sent to Frederick N. Dyer, Experimental Psychology Division, U.S. Army Medical Research Laboratory, Fort Knox, Kentucky 40121.

unique information about the normal processes that are disrupted (e.g., optical illusions), the present study was designed to further explore the delaying effect of these combination color and word stimuli on cross-attribute matching. Such combination stimuli constitute the basic elements of the Stroop (1935) test, where the time for color naming with a series of colored color names (each of which is printed in a color that does not correspond to the color name) is delayed in comparison to the time for color naming with a series of rectangular color patches. The Stroop test itself provides a means for studying the interaction of central processes associated with color and word perception (Dyer, in press). The Treisman and Fearnley (1969) procedure also lends itself to such investigations and, in fact, has an advantage in that only the two responses—same and different—are required regardless of the number of levels of the perceptual dimension.

In addition, the Treisman and Fearnley (1969) matching paradigm is excellently suited for the study of the processing of the same and different judgments themselves. Their results indicated that these judgments were disrupted for cross-attribute matching with irrelevant stimuli present. If different processes are involved in same and different judgments as has been suggested (Bindra, Donderi, & Nishisato, 1968), these judgments may differentially reflect effects of redundant and/or conflicting relationships between the irrelevant and relevant stimuli.

In the present study, the times for individual verbal same and different responses were analyzed instead of the total card-sorting times of Treisman and Fearnley (1969). This allowed separate comparisons of matching latencies for the following seven categories of relationship among the two relevant stimuli and the irrelevant stimulus:

Relevant values matching
(same judgments)

Control–Same—no irrelevant stimulus, upper stimulus is a word (for WC matches) or color (for CW matches) but not both;

Congruent–Same—upper irrelevant stimulus, color or word, corresponds to the relevant stimuli;

Incongruent–Same—upper irrelevant stimulus is different from the relevant stimuli.

Relevant values nonmatching
(different judgments)

Control–Different—no irrelevant stimulus, upper stimulus is either a word or color but not both;

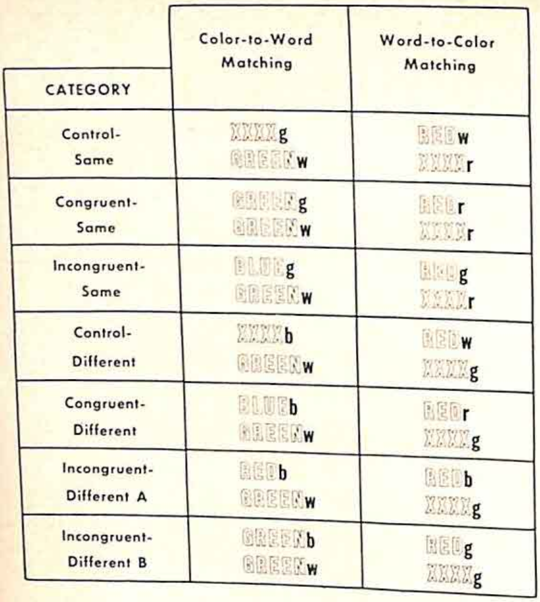
Congruent–Different—upper irrelevant stimulus differs from lower stimulus but agrees with the relevant stimulus it is combined with;

Incongruent–Different A—upper irrelevant stimulus differs from both relevant stimuli;

Incongruent–Different B—upper irrelevant stimulus agrees with lower relevant stimulus (which implies a disagreement with the combined upper relevant stimulus).

Figure 1 presents a pair of stimulus configurations for each of these seven categories. The first example of each pair is for CW matching and the other is for WC matching. It is suggested that the outline letters in Figure 1 be colored to clarify these easily confusable relationships. The lowercase letters r, b, g, and w stand for the colors red, blue, green, and white. White is neutral for these stimuli as black was for the stimuli of Treisman and Fearnley (1969).

The various redundant and conflicting relationships between the irrelevant stimulus and the relevant stimuli suggest some predictions of differences in time for same and different judgments of the seven categories. For the Congruent–Same, Congruent–Different, and Incongruent–Different A categories, the relationship of the irrelevant aspect of the upper combination stimulus to the lower stimulus is the same as the critical relationship of the relevant aspect of the upper stimulus to the lower



CATEGORY	Color-to-Word Matching	Word-to-Color Matching
Control-Same	XXXX$_g$ GREEN$_w$	RED$_w$ XXXX$_r$
Congruent-Same	GREEN$_g$ GREEN$_w$	RED$_r$ XXXX$_r$
Incongruent-Same	BLUE$_g$ GREEN$_w$	RED$_g$ XXXX$_r$
Control-Different	XXXX$_b$ GREEN$_w$	RED$_w$ XXXX$_g$
Congruent-Different	BLUE$_b$ GREEN$_w$	RED$_r$ XXXX$_g$
Incongruent-Different A	RED$_b$ GREEN$_w$	RED$_b$ XXXX$_g$
Incongruent-Different B	GREEN$_b$ GREEN$_w$	RED$_g$ XXXX$_g$

FIG. 1. Examples of stimulus pairs for different categories of relationship between irrelevant and relevant aspects of the stimuli. (Lowercase subscripts denote color—red, blue, green, or white.)

stimulus. This redundancy might be expected to permit fast judgments of same (Congruent–Same) or different (Congruent–Different and Incongruent–Different A) for these categories. On the other hand, this relationship conflicts for the Incongruent–Same and Incongruent–Different B categories, and classification of these stimuli might be expected to take longer.

METHOD

Subjects. Sixteen males ranging in age 20–24 yr. who had recently completed Army basic training served as *S*s in the experiment. All had adequate visual acuity, with correction if necessary, and color vision within normal limits as measured by the Bausch and Lomb Orthorater.

Apparatus. Twelve 35-mm. slides were constructed for each of the seven categories. One half were colors combined with irrelevant words above white words, and the other half were words combined with irrelevant colors above colored Xs for CW and WC matches, respectively (see Figure 1). The stimuli represented all possible combinations of the three colors (red, blue, and green) and the three words (*red*, *blue*, and *green*). For the Incongruent–Same and all of the Different categories, the six slides for a category and match direction were all different. For the Control-Same and Congruent–Same categories, only three different slides are possible for each match direction. To make equal numbers of stimuli in all categories, each of these slides was used twice.

Slides were projected at a distance of 150 cm. The background of the words was dark. The height of the projected capital letters subtended a visual angle of about 2°, and the separation between the two words (or the word and the string of Xs) was equal to the letter height.

Verbal response latencies were measured by a voice-operated switch which stopped a Hunter Klockounter that had started with presentation of the stimulus.

Procedure. The *S*s were tested at two sessions with about a week between them. Within each session, two random sequences of 42 stimuli were presented. Stimuli for WC matching were in one sequence and stimuli for CW matching in the other. Six stimuli from each of the seven categories were included in a sequence. Half of the *S*s began with WC matching and half with CW matching, and at the second session this order was reversed.

Every random sequence of 42 stimuli in the experiment was different. Prior to each sequence, examples of the stimuli were presented and instructions for the required type of matching were given. At the first session, practice trials were given prior to each sequence. The interval between stimuli in a sequence was 10 sec. The *S* rested for 2 min. between sequences. Instructions stressed both speed and accuracy. Errors were noted and such trials were repeated in a short sequence of additional stimuli at the end of each regular sequence.

RESULTS

Response latencies for seven general categories of relevant and irrelevant stimulus combinations were analyzed as one factor (category) in a repeated-measures analysis of variance that also included a two-level factor of match directions, another two-level factor of sessions, and a six-level factor (combination) that represented the different particular combinations of the three colors (red, blue, and green) and the three words (*red*, *blue*, and *green*) used as attribute values. The standard deviation of the responses latencies, calculated from the nearly equal error terms for three- and four-way interactions, was 204 msec.

Highly significant differences appeared for the seven categories, $F\ (6,\ 90) = 39.4$, $p < .001$. The mean response latency for the Control–Same category was 914 msec. The Congruent–Same category, in which the irrelevant and relevant stimuli coincided, produced an average response latency of 859 msec. A Newman-Keuls test of

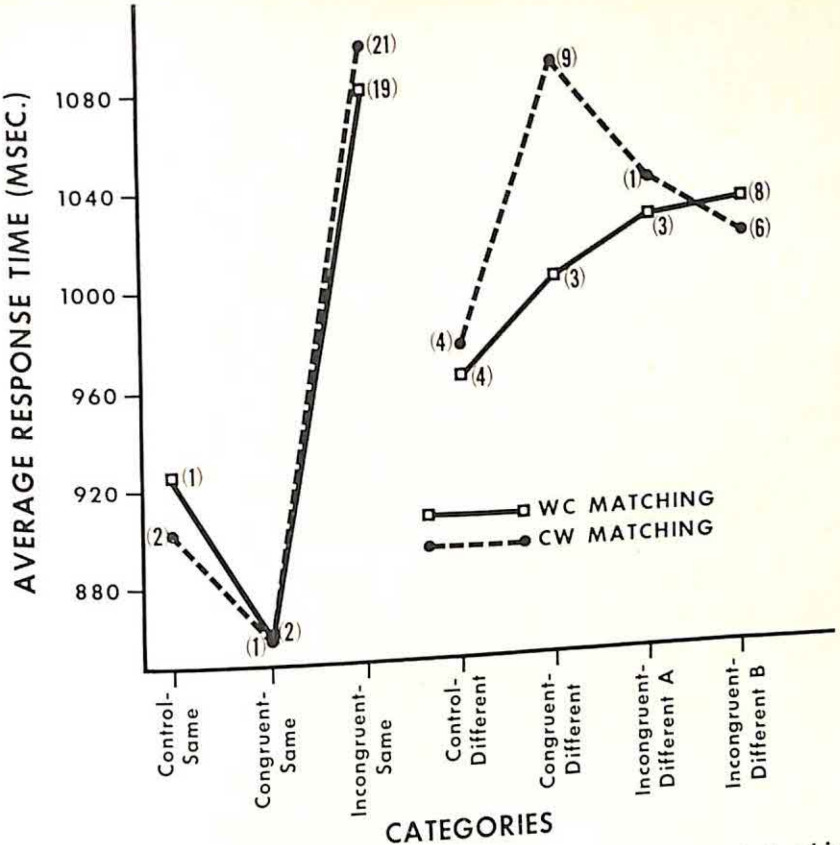


FIG. 2. Average response times for different categories of relationship between stimuli for the two directions of match. (Numbers in parentheses following data points are total errors for that condition.)

ordered means indicated this to be significantly faster ($p < .01$) than the Control–Same category. The Incongruent–Same category, where the irrelevant stimulus conflicted with the matching relevant stimuli, produced an average response latency of 1,093 msec. This response latency was significantly longer ($p < .01$) than that of the six other categories.

Among the four categories requiring different responses, the Control–Different category with no irrelevant stimulus produced an average response latency of 971 msec. This was significantly faster than the other Different categories and significantly slower than the Control–Same and Congruent–Same categories ($p < .01$ for all comparisons). The Congruent–Different, Incongruent–Different A, and Incongruent–Different B categories did not differ from each other at the .05 level of significance. Means for these conditions were 1,047 msec., 1,034 msec., and 1,025 msec., respectively. Categories producing long response latencies also produced high error rates (see Figure 2). The rank-order correlation between response latency and error rate for the seven categories was .81.

The match direction main effect did not approach significance and the only interaction with this variable that approached significance was that involving category, $F(6, 90) = 1.83$, $p < .10$. This interaction was significant for a separate analysis of the four categories requiring different responses, $F(3, 45) = 3.06$, $p < .05$. Average response times for the two match directions are plotted for all of the categories in Figure 2. The only appreciable difference for the two directions of match occurred for the Congruent–Different category—CW matching was 87 msec. longer, on the average, than WC matching for this category. The parenthesized number following each plotted point was the total number of errors in the experiment for that condition. The longer response latency for CW matching for the Congruent–Different category was accompanied by a higher error rate with nine incorrect same responses compared to only three for WC matching.

These differences in response latency and errors between the two match directions for the Congruent–Different category provide a strong clue to the nature of central processing of colors and words, at least as it occurred in the present study, and are discussed later in considerable detail.

The six levels of the combination factor corresponding to the particular different combinations of the colors and words represented somewhat different things for three subgroupings of the seven categories. Despite this fact, a significant main effect for combinations appeared for the overall analysis, F (5, 75) = 13.0, $p < .001$. All four Different categories utilized the same criterion to designate these levels, and in the separate analysis of these categories the main effect was also significant, F (5, 75) = 9.0, $p < .001$. The abscissa of Figure 3 designates the relevant word value and relevant color value that combine to differentiate these different levels. For example, at the first point on the abscissa the relevant word is *red* and the relevant color is blue. Main effect means for these six levels of the combination factor fall at the midpoints between the two curves presented in Figure 3. Of the 15 possible individual comparisons between these six means, the Newman-Keuls procedure indicated nine to be significant at at least the .05 level. Designating the points on the abscissa from left to right as Level 1 to Level 6 of the combination factor, Level 1 differed from all other levels with the exception of Level 5. Level 4 differed from Level 5, and Level 6 differed from all other levels.

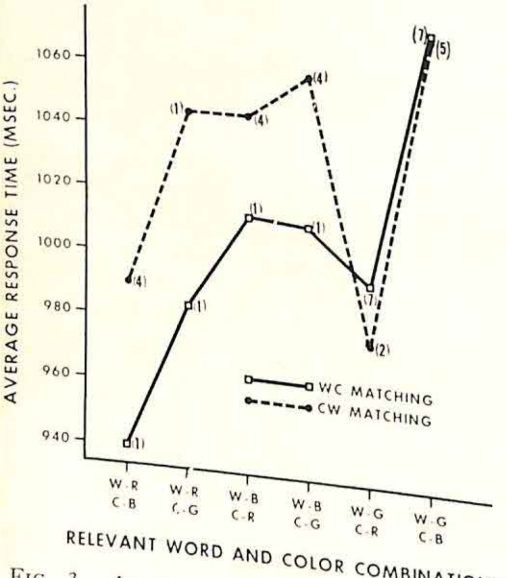


FIG. 3. Average response times for Different categories for particular combinations of words (W) and colors (C). (R = red; B = blue; G = green. Numbers in parentheses following data points are total errors for that condition.)

The interaction appearing in Figure 3 between combination and match direction was not significant, F (5, 75) = 1.55, $p < .20$. It is plotted primarily to illustrate the main effect, but it also indicates possible differences between CW and WC matching for the stimuli containing the words *red* and *blue*. Parenthesized numbers again represent the number of errors in the experiment for the plotted condition.

The main effect of sessions was highly significant, F (1, 15) = 13.9, $p < .01$. Average response time for the first session was 1,037 msec. and for the second, 947 msec. No interaction of other variables with this factor was significant, indicating that this improvement over sessions was fairly constant for different categories, directions of match, and particular word-color combinations. No other main effects or interactions in the analyses approached significance.

DISCUSSION

Basic differences between same and different judgments are illustrated in the present results. The faster time for same judgments in Control–Same and Congruent–Same categories than for different judgments parallels the findings of Bindra et al. (1968) for within-attribute comparisons of colors; Posner, Boies, Eichelman, and Taylor (1969) for within-attribute comparisons of letters; and Seymour (1970) for cross-attribute comparisons between words and shapes. Such faster production of same judgments led Bindra et al. (1968) to conclude that different decision processes operated for same and different judgments. New evidence for separate decision processes comes from the even faster response times for same responses when the irrelevant stimulus agrees with the relevant stimuli (Congruent–Same) than in the Control–Same condition. Congruent combinations of words and colors can actually increase the speed of color naming over control conditions in Stroop paradigms using individual stimuli (Dyer, 1971b), and a possible explanation of the fast same responses in the Congruent–Same category would be that the congruent "irrelevant" word speeds encoding of the color stimulus. However, this would also predict fast different responses for the Congruent–Different category which definitely did not appear. In fact, every category of different judgments was significantly slower than the Control–Different condition. This differential utilization of redundancy in same judgments than in different judgments argues strongly for different decision processes for the two types of response.

Along with this illustration of basic differences between same and different judgments, considerable insight into the processes underlying cross-attribute matching can also be gained from the present data. The primary clue to these processes is the large difference in response time between the two match directions for the Congruent–Different category. The irrelevant stimulus and the relevant stimulus with which it was combined were equivalent in this category. When the match direction was color to word, with the irrelevant stimulus being a word, different responses were greatly delayed relative to the other match direction, in which the irrelevant stimulus was a color (see Figure 2). In addition, error rates were much higher for the color-to-word match direction.

Cross-attribute comparisons require transformation of one or both stimulus attributes to make them comparable at some central location. The difference in the Congruent–Different category for the different match directions can be explained by assuming that less transformation is required for one attribute than the other, and thus the form of the untransformed *irrelevant* stimulus for one match direction is already close to that form in which the central comparison is made. This assumption implies that for the other match direction the irrelevant stimulus would require transformation to be in this "central comparison" form. The situation with the irrelevant stimulus near to the form of the central comparison would be expected to provide more difficulty for the making of comparisons for at least two reasons. First, the relevant stimulus with which the irrelevant stimulus is combined *would* require transformation, and while this additional process is carried out, attention would be directed to the exact position of the irrelevant stimulus. Second, comparisons of the transformed relevant stimulus could easily be erroneously made to the irrelevant stimulus since it is in the same form as the other relevant stimulus involved in the comparison.

Since it was in matches of colors to words, where the irrelevant stimulus was a word, that long response times and high error rates occurred, the implication of the above reasoning is that in comparisons between words and colors the relevant color is transformed into a word or, at least, into a form that is more similar to words than to colors. In addition, this transformation of colors to words is implied for both directions of match.

Given that this analysis of the differences for the Congruent–Different category is correct, some predictions also follow for the other Different categories. For the Incongruent–Different B category, where the irrelevant stimulus is the same as the lower relevant stimulus, an accidental comparison during CW matching of a transformed color to the irrelevant word instead of to the relevant word would not expend as much time as it did in the Incongruent–Different A category because the irrelevant and relevant words are the same. When this similarity was noted—and the result of Treisman and Fearnley (1969) showing fast within-attribute matching suggests this would not take long—the erroneous comparison with the irrelevant word would not have to be repeated. The average response time for the Incongruent–Different B category was 23 msec. faster than for the Incongruent–Different A category (CW matches), and although this was not significant, it is in the direction that supports this interpretation. In addition, the Incongruent–Different B category was the only Different category that allowed faster comparisons for CW matches than for WC matches.

Unlike the conclusion of the present study that the attribute of color is transformed to a word code, Seymour (1970) concluded that in comparisons of words and shapes, words are converted to shape codes for the central match. His conclusion was reached on the basis of faster comparisons when the word was fixated first than when the shape was fixated first. Fixation was not controlled in the present study, but if one assumes that the top of the two stimuli was fixated first, then the different results for the two match directions for the Control–Same category, although not significant, are again contrary to those of Seymour for words and shapes. A basic difference may exist between the processing of color and shape. Further research is planned where comparisons between colors and words are made with the order of fixation controlled.

The differences that appeared for particular combinations of words and colors are not readily explained. Differences in time for processing of particular words and colors may constitute a partial explanation. But the data presented in Figure 3 indicate that color effects depend on the word and that word effects depend on the color. Further explorations utilizing additional colors and words may permit better understanding of these reliable combination differences.

In providing evidence about the form in which the central comparison of colors and words was made, the Congruent–Different category stimuli, in conjunction with stimuli from other categories, demonstrate potential for similar analyses of other stimulus dimensions. Dyer (1971a) explored a form of the Stroop test that utilized achromatic shades and names of achromatic shades in combination instead of colors and color names. Shor (1970) combined arrows with words designating directions and also obtained a small amount of interference similar to that of the Stroop test. Dyer (1972) found a small but significant amount of interference to direction naming with incongruent direction names as the movement stimuli. It should be possible to construct stimulus pairs similar to those of the present study but combining values of other visual stimulus dimensions such as achromatic shade, spatial position, movement direction, size, etc., with words designating values of the respective dimensions. Analysis of the time required for comparisons between the words and stimulus values for different categories and match directions should permit determination of the central transformation that takes place to allow these comparisons.

REFERENCES

Bindra, D., Donderi, D. C., & Nishisato, S. Decision latencies of "same" and "different" judgments. *Perception & Psychophysics*, 1968, **3**, 121–130.

Dyer, F. N. A comparison of chromatic and achromatic versions of the Stroop color-word test. *Psychonomic Science*, 1971, **22**, 235–237. (a)

Dyer, F. N. The duration of word meaning responses: Stroop interference for different preexposures of the word. *Psychonomic Science*, 1971, **25**, 229–231. (b)

Dyer, F. N. Latencies for movement naming with congruent and incongruent word stimuli. *Perception & Psychophysics*, 1972, **11**, 377–380.

Dyer, F. N. The Stroop phenomenon and its use in the study of perceptual, cognitive and response processes. *Memory & Cognition*, 1973, **1**, in press.

Posner, M. I., Boies, S. J., Eichelman, W. H., & Taylor, R. L. Retention of visual and name codes of single letters. *Journal of Experimental Psychology Monograph*, 1969, **79** (1, Pt. 2).

Seymour, P. H. K. Order of fixation effects in classification of word-shape pairs. *Quarterly Journal of Experimental Psychology*, 1970, **22**, 440–449.

Shor, R. E. The processing of conceptual information on spatial directions from pictorial and linguistic symbols. *Acta Psychologica*, 1970, **32**, 346–365.

Stroop, J. R. Studies of interference in serial verbal reactions. *Journal of Experimental Psychology*, 1935, **18**, 643–662.

Treisman, A., & Fearnley, S. The Stroop test: Selective attention to colours and words. *Nature*, 1969, **222**, 437–439.

(Received June 9, 1972)

Journal of Experimental Psychology
1973, Vol. 98, No. 1, 109–112

ACQUISITION OF A BAR-PRESS RESPONSE TO ESCAPE FRUSTRATIVE NONREWARD AND REDUCED REWARD [1]

HELEN B. DALY [2] AND JAMES H. McCROSKERY

State University of New York, College at Oswego

Rats given reinforced placement trials to a 15-pellet reward were subsequently given either 1 pellet or 0 pellets and were lifted out of the box if they pressed a bar. Bar-press speeds were faster for these two experimental groups than for control groups which had not been given the 15-pellet reward. These results indicated that rats would learn a bar-press response to escape frustrative nonreward and reduced reward.

Frustration theory (Amsel, 1958) assumes that nonreward in the presence of cues previously paired with reward arouses an aversive emotional response. This emotional response, called frustration, is presumed to have inhibitory, excitatory, and reinforcing properties. The inhibitory effects are measured through decreases in response magnitude during extinction and have been demonstrated in numerous behavioral situations (e.g., Kimble, 1961). The excitatory effects have been demonstrated most often in the double runway, and the frustration effect—faster running in the second runway following nonreward than following reward in the first runway—has been obtained under many conditions (e.g., Amsel, 1958; Daly, 1968; Wagner, 1959). The energizing properties of frustration have also been measured by increases in intensity of a bar-press response following extinction of a chain-pull response (Levine & Loesch, 1967) and by increases in the magnitude of the startle response (Wagner, 1963). The reinforcing effects of frustration have been shown only by the demonstration that rats will acquire a hurdle-jump response to escape frustration (e.g., Daly, 1969a, 1969b). To give additional support to the assertion that frustration reduction can reinforce any instrumental escape response, acquisition of responses other than hurdle jumping must be demonstrated. One purpose of the present experiments was to determine if rats would learn to press a bar to escape frustration aroused by nonreward. A second purpose was to determine if rats would learn a bar-press response to escape frustration aroused by reduced reward, since reductions in reward magnitude have also been shown to arouse frustration (e.g., Daly, 1968, 1969b).

EXPERIMENT I

Method

Subjects. Sixty male Holtzman rats obtained from Holtzman Co., Madison, Wisconsin, 90–95 days old and weighing 315–375 gm. on the first day of deprivation, were reduced to and maintained at 80% of their original body weight. They were housed in individual cages throughout the experiment.

Apparatus. A black painted box, 30 × 12 × 17 cm., contained a bent measuring spoon, hanging on the long wall 9 cm. from the end and 5 cm. from the floor, in which *S*s received the designated number of 37-mg. pellets (P. J. Noyes Co., Formula A). A response lever (Lehigh Valley Electronics, Inc., Model 1535), mounted 4 cm. from the floor between the measuring spoon and the wall, was interchangeable with a piece of wood which fit flush with the wall. Placement of *S* into the box broke a photobeam 14 cm. from the floor, which started a Hunter Klockounter (.01 sec.). When *S* pressed the lever, the Klockounter was stopped. A 6-v. light, 10 cm. from the floor between the measuring spoon and the lever, was turned on when the photobeam was broken and turned off by a bar-press response.

Procedure. Adjustment to 80% body weight took 8 days. For the next 10 days *S*s were placed six times per day into the box with the food cup present, but the light was not turned on and the bar was not present. There were two experimental groups and two control groups with 15 *S*s per group: Experimental Groups 15-0 and 15-1 received 15 pellets in the food cup during the training phase and remained in the box until finished eating; Control Groups 0-0 and 1-1 received no pellets and one pellet, respectively, in the food cup during the training phase and remained in the box for 10 sec. Following each placement trial *S* was returned to a retaining cage where water was available. For the next 5 days (test phase), the light came on as *Ss* were placed in the box and the bar was present. Nine trials were given on each of the first 4 days and six trials were given on the fifth day. Groups 15-0 and 0-0 received no pellets and Groups 15-1 and 1-1 received one pellet in the food cup. If *S* pressed the bar, the light was turned off and *S* was removed from the box and returned to the retaining cage. If *S* did not press the bar within 60 sec., he was removed and returned to the retaining cage and a latency of 60 sec. was recorded. The *Ss* were run in squads of four, one from each group. Since *Ss* were run in rotation, the intertrial interval varied from trial to trial due to variations in eating time and bar-press latency but ranged between 2–4 min. Two *Es* ran some of the *Ss*. However, both *Es* handled *Ss* prior to the training phase and no *E* ran an *S* for more than 2 days in a row.

---

[1] This research was supported by National Institute of Mental Health Grant 17398 awarded to the first author. The authors wish to thank Lynn Tondat for running a portion of the animals and doing the statistical analyses.

[2] Requests for reprints should be sent to Helen B. Daly, Department of Psychology, State University of New York, Oswego, New York 13126.

Results

Mean bar-press speeds (1/sec) in blocks of three trials are presented in Figure 1. In general, the two experimental groups, Groups 15-0 and 15-1, pressed the bar faster than the two control groups, Groups 0-0 and 1-1.

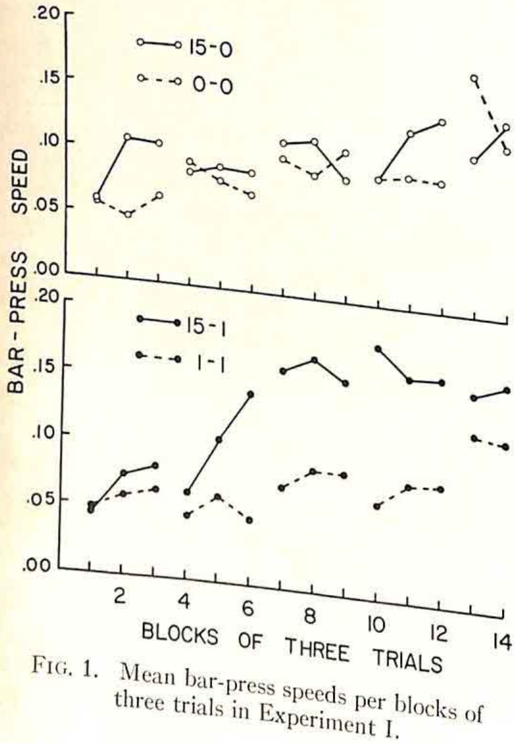


FIG. 1. Mean bar-press speeds per blocks of three trials in Experiment I.

Results of an analysis of variance (2 × 2 × 14) indicated that the experimental groups pressed the bar faster than the control groups, F (1, 56) = 4.82, $p < .05$; that bar-press speeds increased across trial blocks, F (13, 728) = 10.52, $p < .001$; and that the Experimental–Control × Trial Blocks interaction effect was significant, F (13, 728) = 1.89, $p < .05$. There was no effect due to one or zero pellets ($F < 1$), and this variable did not interact with any other (all ps > .05). To determine if all groups showed increases in bar-press speeds across trial blocks, comparisons between Trial Blocks 1 and 14 were made. Groups 15-0, 15-1, 0-0, and 1-1 all showed significant increases in bar-press speeds, t (14) = 3.06, 4.28, 3.06, and 3.23, respectively, all ps < .01.

Discussion

Both the extinction and reduced-reward experimental conditions resulted in faster bar-press speeds than the two control conditions. Presumably, *Ss* in the experimental groups had an expectancy (r_r) for the large reward conditioned during the acquisition phase, and the small reward or empty cup aroused frustration during the test phase. Reduction in frustration when *S* was removed from the box immediately after the bar press reinforced the response.

One problem exists with the interpretation of the results due to the finding that Control Groups 1-1 and 0-0 showed acquisition of the bar-press response. To conclude that frustration reduction can reinforce an instrumental response, one must eliminate all other sources of reinforcement because previous investigations have demonstrated increases in response strength due to an irrelevant drive (e.g., Ley, 1965). It could be that confinement in the box was aversive since control *Ss* were left in the box for 10 sec. during the first 60 trials and up to 60 sec. if they did not press the bar during the test phase. Therefore, control *Ss* may have learned to escape confinement by pressing the bar. It is also possible that control *Ss* pressed the bar to turn off the light. It has been shown that rats will learn to press a bar to decrease illumination in the chamber (e.g., Goodrick, 1970; Lockard, 1966). The purpose of the light was to increase the amount of frustration in Groups 15-0 and 15-1, since a previous study (Daly, 1969b) had shown that groups given frustration conditioning to apparatus cues plus a light jumped a hurdle faster to escape these cues than if frustration was conditioned only to the apparatus cues. In that study the light had no measurable effect on the control group's behavior. In the present study it may have. Therefore, either escape from confinement and/or the decrease in illumination could have reinforced the bar-press response. Frustration may have merely been an irrelevant drive which energized responding in the experimental groups by adding to the total drive level, rather than providing reinforcement for the bar-press response via its reduction following removal from the box. The purposes of the second experiment were (*a*) to replicate the extinction condition (Groups 15-0 and 0-0) of Experiment I, since the differences between these groups were not very convincing despite the statistical significance of the results; and (*b*) to determine if the control group would show acquisition of the bar-press response if more time was spent in the box during training and if no light was present during the bar-press test phase.

Experiment II

Method

Subjects and apparatus. Thirty male Holtzman rats obtained from Holtzman Co., Madison, Wisconsin, 95–100 days old and weighing 330–400 gm. on the first day of deprivation, were reduced to and maintained at 80% of their original body weight. The *S*s were housed in individual cages. The apparatus from Experiment I was used, but the light was not present.

Procedure. The differences in procedure from Experiment I were as follows: (*a*) All *S*s were placed 15 times per day into the box during the training phase (5 days) and test phase (2 days) of the experiment; (*b*) both Groups 15-0 and 0-0 (15 *S*s per group) remained in the box for 60 sec. in the training phase; (*c*) *S*s were run in squads of two, one from each group; and (*d*) the intertrial interval was 1 min. during both phases.

Results

Mean bar-press speeds (1/sec) in blocks of three trials are presented in Figure 2. Group 15-0 clearly learned to press the bar faster than Group 0-0, which showed no increase in bar-press speeds across trial blocks. Results of an analysis of variance (2 × 10) indicated a significant Groups × Trial Blocks interaction, F (9, 252) = 1.95, $p < .05$. Tests of simple main effects indicated that there were no differences between the groups on Trial Blocks 1, 2, 3, 6, 7, 8, and 10 (all ps $> .05$), but that Group 15-0 pressed the bar faster than Group 0-0 on Trial Blocks 4, 5, and 9, F (1, 28) = 6.26, 5.60, and 4.76, respectively, all ps $< .05$. Results of t tests between Trial Blocks 1 and 5 indicated a significant increase in bar-press speeds for Group 15-0, t (14) = 2.21, $p < .05$; and a significant decrease in bar-press speeds for Group 0-0, t (14) = 2.88, $p < .02$.

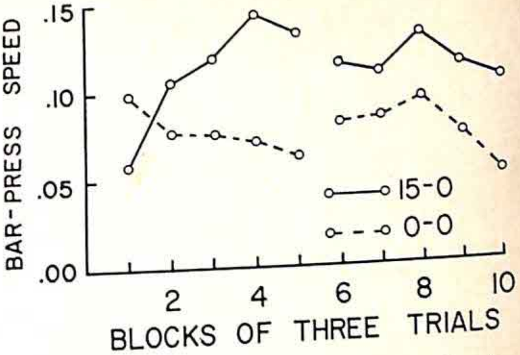


FIG. 2. Mean bar-press speeds per blocks of three trials in Experiment II.

Discussion

The procedures used in Experiment II resulted in an increase in bar-press speeds in Group 15-0 and a decrease in bar-press speeds in Group 0-0. These data provide support for the frustration analysis. Nonreward during the bar-press phase for *S*s previously reinforced presumably aroused frustration. Removal from the box, which was contingent on bar pressing, reduced frustration which reinforced the bar-press response. In both experiments the differences between the experimental and control groups increased early in testing and then tended to decrease with additional trials. This result also occurred in the previously cited hurdle-jump studies (Daly, 1969a, 1969b) and is predicted by frustration theory. Reward expectancy is conditioned during training. When the reward magnitude is reduced to zero or nonzero levels, frustration is aroused and presumably extinction of the expectancy for the large reward also occurs. As the expectancy for the large reward decreases, the magnitude of frustration aroused should also decrease. Since bar-press responding is based on frustration, one would expect a decrease in responding with extended testing as the expectancy for reward extinguishes and the resultant magnitude of frustration aroused decreases.

In Experiment II, confinement in the box with no light present was not aversive, since Group 0-0 showed a decrease in bar-press speeds. Although it is unclear which of the procedural changes caused the decrease rather than the increase in bar-press speeds in Group 0-0, the most likely reasons involve the changes from 10- to 60-sec. confinement during training and/or the absence of the light in the test phase.

The fact that the frustrated group learned the bar-press response in a situation in which the control group did not is theoretically important. In Experiment I, either escape from confinement or the decrease in illumination could have reinforced the bar-press response in both experimental and control groups, and as mentioned previously, frustration may have been an irrelevant drive which energized responding of the experimental groups by adding to the total drive level. However, in Experiment II there was presumably no other reinforcement present since the control group did not acquire the response. The only difference between the experimental and control groups was the procedure which aroused frustration. Thus, one can conclude that frustration reduction reinforced the bar-press response.

The present results are especially interesting since Bolles (1970) has argued that acquisition of a bar-press response to escape or avoid aversive stimuli is extremely difficult. Although the rapid learning demonstrated in these experiments seems to run counter to the general finding that acquisition of a bar-press response to escape aversive stimulation is difficult, the present procedure made the bar-press response effective in escaping the entire situation (flight), in contrast to the typical bar-press avoidance or escape situation in which *S* remains in the box and merely escapes or avoids the noxious stimulus. Perhaps this procedural difference made acquisition of what Bolles calls a non-species-specific defense reaction possible.

Previous experiments (Daly, 1969a, 1969b) have shown acquisition of a hurdle-jump response to escape from a frustration-arousing situation. The data from the present experiments demonstrate that a bar-press response can also be learned to escape from frustration. The data thus provide additional evidence that frustration reduction can reinforce instrumental escape responses, and they give further support to the statement that frustration reduction can reinforce any instrumental escape response.

REFERENCES

Amsel, A. The role of frustrative nonreward in noncontinuous reward situations. *Psychological Bulletin*, 1958, **55**, 102–119.

Bolles, R. C. Species-specific defense reactions and avoidance learning. *Psychological Review*, 1970, **77**, 32–48.

Daly, H. B. Excitatory and inhibitory effects of complete and incomplete reward reduction in the double runway. *Journal of Experimental Psychology*, 1968, **76**, 430–438.

Daly, H. B. Is responding necessary for nonreward following reward to be frustrating? *Journal of Experimental Psychology*, 1969, **80**, 186–187. (a)

Daly, H. B. Learning of a hurdle-jump response to escape cues paired with reduced reward or frustrative nonreward. *Journal of Experimental Psychology*, 1969, **79**, 146–157. (b)

Goodrick, C. Light- and dark-contingent bar pressing in the rat as a function of age and motivation. *Journal of Comparative and Physiological Psychology*, 1970, **73**, 100–104.

Kimble, G. A. *Hilgard and Marquis' conditioning and learning*. New York: Appleton-Century-Crofts, 1961.

Levine, G., & Loesch, R. Generality of response intensity following nonreinforcement. *Journal of Experimental Psychology*, 1967, **75**, 97–102.

Ley, R. Effects of food and water deprivation on the performance of a response motivated by acquired fear. *Journal of Experimental Psychology*, 1965, **69**, 583–589.

Lockard, R. B. Several tests of stimulus-change and preference theory in relation to light-controlled behavior of rats. *Journal of Comparative and Physiological Psychology*, 1966, **62**, 415–426.

Wagner, A. R. The role of reinforcement and nonreinforcement in an "apparent frustration effect." *Journal of Experimental Psychology*, 1959, **57**, 130–136.

Wagner, A. R. Conditioned frustration as a learned drive. *Journal of Experimental Psychology*, 1963, **66**, 142–148.

(Received June 12, 1972)

Journal of Experimental Psychology
1973, Vol. 98, No. 1, 113–118

LOSS OF SPATIAL AND IDENTITY INFORMATION FOLLOWING A TACHISTOSCOPIC EXPOSURE[1]

V. M. TOWNSEND[2]

Queen's University at Kingston, Ontario, Canada

Loss of information following a tachistoscopic exposure is examined in three partial-report experiments. The first is essentially a replication of a 1961 experiment by E. Averbach and A. S. Coriell. Analysis of the type of errors *S*s made revealed that identity information, i.e., the names of the elements in a display, is not lost as the partial-report cue is delayed. The results of the second experiment (a recognition experiment) supported the finding. A third experiment demonstrated that *S*s do not lose the spatial location of the probe relative to the stimulus display as a whole. It is suggested that the decay function found in partial-report experiments using a visual probe may result from *S*s' inability to retain information about where a letter is relative to a visual probe.

Sperling (1960) and Averbach and Coriell (1961), using two different methods of partial report, have shown that more information is available for report immediately following a tachistoscopic exposure than can be retained in short-term memory. Their experiments have led to a widely accepted general model which holds that a visual stimulus is initially registered in iconic memory and that some of the stimulus information is then transferred to short-term memory (Averbach & Sperling, 1961; Mackworth, 1963; Neisser, 1967; Sperling, 1963, 1967). Although other techniques have provided data consistent with the model (e.g., Eriksen & Collins, 1967; Haber & Standing, 1969; Mewhort, Merikle, & Bryden, 1969; Smith & Carey, 1966), the primary evidence for it has been obtained using partial-report sampling techniques.

In partial-report experiments, *S* is shown an array and is cued to report part of that array. Dick (1967, 1969), however, noticed that the classic partial-report experiments have involved the use of a spatial cue, i.e., the part to be reported is indicated via its spatial location. Thus, in order to report correctly the identity of probed elements, *S* must also retain the spatial location of those elements. Using a partial-report procedure similar to that of Sperling (1960), Dick examined the loss of different kinds of information from iconic memory. He concluded that different kinds of information are subject to decay at different rates and argued that identity information, the names of the elements in a display, is not lost as rapidly from the icon as spatial information.

The three experiments presented here follow from Dick's (1967, 1969) work and were designed to provide a more stringent test of the notion that spatial and identity information are lost independently.

EXPERIMENT I

The first experiment is essentially a replication of Averbach and Coriell (1961). A row of eight random letters was exposed on a tachistoscope, and *S* was required to report one letter from the display. A bar marker was presented after the exposure to indicate the position of the letter to be reported. As in the Averbach and Coriell study, cue delay was systematically varied, but in the present study the duration of the cue was also varied. Both a long cue and a short cue were used as a measure of whether the decay function normally attributed to loss of information from iconic memory also reflects cue uncertainty. If spatial information is lost more rapidly than identity information, perhaps the brief cue used by Averbach and Coriell would not give *S* sufficient time to identify and localize its spatial location relative to the elements in the array. Therefore, accuracy, as measured by the correct identification of the elements in the display, may have been confounded by cue uncertainty about the locus of the cue.

---

[1] The research reported here is based on an unpublished master's thesis and was supported by National Research Council of Canada Grant APA-318 to D. J. K. Mewhort. The author wishes to thank Pat Crawford for her assistance in replicating the experiments.

[2] Requests for reprints should be sent to V. M. Townsend, who is now at the Psychological Laboratory, University of Cambridge, Downing Street, Cambridge, England.

Method

Subjects. Sixteen undergraduates enrolled in an introductory course in psychology at Queen's University served as *S*s in the experiment.

Apparatus and procedure. Eight-item random letter sequences were printed on white cards with Letraset Instant Lettering (No. 287) in uppercase letters. No letter appeared more than once per sequence. The visual substance of the display was approximately $4° \times \frac{1}{2}°$. Cues were printed on similar cards using the letter I of the Letraset Lettering. One cue was prepared to correspond to the position of each of the eight letters presented, and all were printed so that one I appeared $\frac{1}{2}°$ above and one I $\frac{1}{2}°$ below the corresponding letter.

The materials were presented binocularly in a Gerbrand's three-field tachistoscope. The luminance of all fields was approximately 4.5 ftl. The stimulus materials were presented for 50 msec. and were followed after a dark interval of 0, 150, 300, or 450 msec., by a 40- or 900-msec. cue. The stimulus was preceded by a centered fixation dot. After the cue had been presented, all fields were dark. The *S* was required to report the letter indicated by the cue and was instructed to guess if he was uncertain as to the identity of the cued letter. The *E* started each trial by presenting the fixation dot. When ready, *S* initiated the presentation and cue sequence.

Each *S* was tested 64 times. The conditions were defined by factorially combining three independent variables: delay of cue (0, 150, 300, 450 msec.), duration of cue (40 or 900 msec.), and serial position in the stimulus (1–8). Both the pairing of the stimulus cards to the test condition and the order of presentation of the three independent variables were independently randomized for each *S*.

Results and Discussion

The proportion of letters correctly identified as a function of the duration and of the delay of the postexposure cue is presented in Figure 1. An analysis of variance performed on the number of correct identifications of the probed letter revealed the following significant effects: serial position, $F(7, 105) = 31.5$, $p < .0001$; cue duration, $F(1, 15) = 8.265$, $p < .02$; cue delay, $F(3, 45) = 3.462$, $p < .02$; and the Duration × Delay interaction, $F(3, 45) = 6.187$, $p < .002$. When the partial-report cue was short (40 msec.), delaying the cue lowered accuracy of report. This result is in accord with Averbach and Coriell's (1961) results. However, when the cue was long (900 msec.), accuracy was low at all delays of the cue. Thus, if the cue is brief, *S* is able to take advantage of iconic memory, but if it is not, he cannot make use of iconic representation. The results for the long cue are not consistent with the expectation that the long cue would help retain spatial location information, thus enabling *S* to encode only identity information. Contrary to the predictions, the longer cue appears to interfere with *S*'s ability to utilize the iconic image. The main effect of serial position was statistically significant and exemplified the W-shaped curve usually found in probe experiments of this sort (Merikle, Lowe, & Coltheart, 1971).

An analysis of the type of errors made by *S*s was undertaken to indicate what information in the icon was disrupted. There are two kinds of errors *S* could make: He could report another letter from the stimulus or he could report one which was not in the stimulus (an intrusion error). Table 1 presents both the number and proportion of intrusion errors, as a function of the duration and of the delay of the postexposure cue. Each cell in the table is based on 128 trials (16 *S*s × 8 Serial Positions). With a mean number of intrusion errors of only 14.8, it is clear the intrusions are rare.

If *S*s could not recognize any letters in the display because of a general masking effect of the sort implied by Eriksen's (1966) luminance-summation theory, the proportion of intrusions should be close to .72 and the proportion of correct responses should be close to chance ($1/25 = .04$). As is shown in the figure, the overall proportion of correct responses is approximately .34. Alternately, if only the probed letter was masked, the proportion of in-

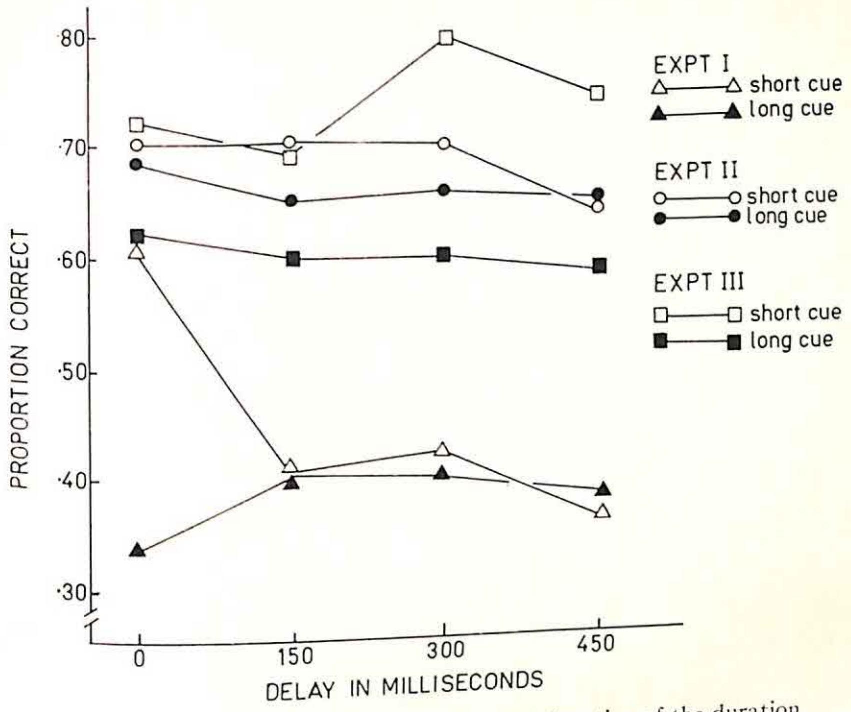


FIG. 1. Proportion of correct responses as a function of the duration and the delay of the postexposure cue.

trusions should be close to .0 and that of correct response should be close to chance (1/19 = .05). Thus, it seems that the bulk of errors occurring in the experiment do not represent failures to see the letters in the display. They might represent, however, failures to keep track of the spatial location of the letters in the display. If this is the case, the delay-of-cue effect, i.e., the decrease in identification accuracy with delay of the postexposure cue, may result because the spatial information fades more rapidly from the icon than the identity information, and identification accuracy as measured by the partial-report technique decreases as *S* loses the spatial location of the letter relative to the bar probe.

TABLE 1
NUMBER AND PROPORTION OF INTRUSION ERRORS (IN PARENTHESES)

Duration of cue (in msec.)	Delay of cue (in msec.)				
	0	150	300	450	*M*
40	14 (.28)	10 (.13)	15 (.21)	20 (.24)	14.8 (.215)
900	14 (.16)	16 (.21)	13 (.17)	16 (.20)	14.8 (.185)
M	14 (.22)	13 (.17)	14 (.19)	18 (.22)	

The following experiments were designed to investigate directly the loss of identity information and spatial locus information from iconic memory. Experiment II concerns the loss of identity information while Experiment III concerns the loss of relative spatial information, i.e., the loss of the spatial location of the probe.

EXPERIMENT II

The stimuli were exposed as in the first experiment, but *S* was cued with a letter instead of a bar marker. The *S*'s task was to indicate whether or not the probed letter was in the stimulus display. On one half of the trials the probed letter was present in the display; on the other half it was absent. In the experiment, the retention of spatial information about the probe and the letter array was minimal, and therefore, the task was one of almost pure identity information.

Method

Subjects. Sixteen undergraduates enrolled in an introductory course in psychology at Queen's University served as *S*s in the experiment.

Apparatus and procedure. The stimulus cards used in this study were identical to those used in the first study. Twenty-six letter cues (uppercase) were printed on white cards with Letraset (No. 287), one letter per card. These were printed so that each letter would appear $\frac{1}{2}$° above the stimulus sequences, between Positions 4 and 5.

The procedure for this experiment was similar to that used in Experiment I. However, in the following study, *S* was required to report whether or not the probed letter was present in the display. The trials on which the probed letter was not in the display were randomly interspersed among the trials on which the probed letter was in the display.

Results and Discussion

The proportion of correct responses as a function of the duration and the delay of the postexposure cue is presented in Figure 1. An analysis of variance on the number of correct responses, given that the probed letter was present in the stimulus, yielded an overall effect of serial position, $F(7, 105) = 5.705$, $p < .001$. No other main effects or interactions were significant.

Whether the postexposure cue was of 40- or 900-msec. duration, delaying the cue did not lower accuracy of report. In fact, accuracy is above chance (50%) for all delays and both durations of cue. Thus, the 900-msec. probe field was not degrading the stimulus in a general way (cf. Eriksen, 1966). In Experiment I, when discussing intrusion errors, it was also suggested that regardless of the duration or the delay of the postexposure cue, *S*s recognized the letters present in the stimulus display. The results of this experiment appear to support that notion. It therefore appears that there is no loss of identity information with a delay of the postexposure cue.

With the exception of serial position, an analysis of the correct responses for the conditions on which the probed letter was not present in the display yielded no significant main effects or interactions. The serial position curves were once again W shaped, although they were much flatter than the serial position curve for Experiment I.

From the analysis of Experiments I and II, it seems that regardless of the duration or the delay of the postexposure cue, *S*s can retain the identity information given in the stimulus display. Identity information does not appear to decay with the delay of the cue. Thus, the drop in performance with cue delay (Averbach & Coriell's, 1961, result) must be explained without involving the idea that material is lost from the icon. Dick's (1967, 1969) suggestion that spatial and identity information are independent seems a reasonable possibility, i.e., the decay curve may result because *S* has lost the spatial location of the probe relative to the elements in the display.

EXPERIMENT III

The following experiment was designed to investigate *S*'s ability to locate the position of the bar marker and to investigate the possibility that during the delay period between the stimulus and probe, extraneous factors such as eye movements had not caused *S*'s frame of reference to shift.

The stimuli were exposed as in the first two experiments. A bar-marker cue, indicating which serial position to report, was presented after the exposure. The *S*'s task was to indicate the serial position (1–8) that the bar demarcated. Note that letter identity and position of letters was irrelevant in this task. Letters were presented to make the visual characteristics of the experiment the same as those of Experiment I.

Method

Subjects. Sixteen undergraduates enrolled in an introductory psychology course at Queen's University served as *S*s in the experiment.

Apparatus and procedure. The stimuli used were the same as Experiment I. In the following study, however, *S* was required to give the serial position that the bar probe demarcated. A card with eight boxes demarcating the serial positions from left to right was taped on the tachistoscope in front of *S* for reference. Therefore, to do the task *S* did not have to count serial positions but could make an absolute judgment.

Results and Discussion

The proportion of correct responses as a function of the delay and the duration of the postexposure cue is presented in Figure 1. An analysis of variance on the serial positions correctly identified yielded two significant main effects: serial position, $F(7, 105) = 3.386$, $p < .003$; and cue duration, $F(1, 15) = 5.501$, $p < .03$. All other main effects and interactions did not reach an acceptable level of significance; in particular there was no significant delay-of-cue effect, indicating that *S* could retain the spatial location of the probe relative to the display during the delay.

There was a significant duration-of-cue effect, accuracy being lower for the 900-msec. cue duration. The duration-of-cue effect was also found in Experiment I where the partial-report cue was also a bar marker. It does appear that there is a peculiar phenomenon operating in the case of the bar-marker probe. Since it was not necessary for *S* to retain any identity information in the present experiment, but to retain only spatial location information of the probe, it appears that the long bar probe may interfere with the processing of spatial information.

The serial position curve for this experiment was U shaped rather than W shaped. Positions 2–7 were identified with about equal accuracy (20%–30% correct identifications) and Serial Positions 1 and 8 at 60% and 70% correct, respectively. Thus, performance on an experiment of the Averbach and Coriell (1961) variety appears to be limited by a symmetrical factor related to *S*'s ability to locate the probe in space. Such a limitation questions the force of recent arguments concerning serial processing based on partial-report studies (e.g., Merikle et al., 1971; Mewhort & Cornett, 1972).

GENERAL DISCUSSION

The first experiment has shown that Averbach and Coriell's (1961) results are not replicable when the postexposure cue is of a long duration. Introducing a cue of long duration interferes with rather than enhances *S*'s ability to use the icon. While it might be argued that the long probe is simply masking the icon, there are several factors in the results which show that masking alone could not explain the effect of the long probe; for example, if the long probe field was degrading the icon in a general way (especially at an interstimulus interval [ISI] of 0), then the proportion of intrusion errors in Experiment I should have been greater for the long cue than for the short cue. This was not the case; the proportions were the same for both. Similarly, in Experiment II, the long cue should have yielded lower identification accuracy than the short cue. Accuracy was, in fact, the same for both cue durations. There was a significant effect of cue duration in Experiment III. This effect is irrelevant to the present argument however; since *S* did not have to identify any elements in the stimulus display, it should not matter that they were degraded in terms of identity: *S* should have been able to identify the spatial location of the probe. Although the proportion of correct identifications of the probe location is lower for the long probe, accuracy is still well above chance ($\frac{1}{8} = .125$). Finally, the results of these experiments have been replicated using a light-light-light cycle. If, for example, luminance-summation masking were degrading the icon, one would not expect replication with light ISIs. Thus, it does appear that masking phenomena cannot explain the long cue effect.

The only evidence of a decay function found in the experiments involved the short cue. It may be that *S* continues to process the cue until it is terminated. Because the icon is decaying during this time, 900 msec. will elapse before *S* can extract any useful information from the icon. Eriksen and Collins (1969) have suggested that it takes about 200 msec. to process an indicator and to direct attention. Therefore, at an ISI of 0, at least 200 msec. may elapse before *S* can make an accurate response. Since the life of an icon is about 250 msec. (Engel, 1970), accuracy will be high at an ISI of 0 (because the icon may still be present) and will decrease as the ISI increases. If *S* continues to process a probe until it is terminated, accuracy would then be lower for a probe of long duration.

The main purpose of the experiments was to examine the independence of spatial and identity information. Experiment II has shown that when identity information alone was probed, there was no loss of this kind of information. The analysis of the types of errors made in the Averbach and Coriell (1961) experiment also indicated that identity information was not lost as the partial-report cue was delayed. Thus, it appears that Dick's

(1967, 1969) hypothesis received some support. However, to infer the loss of one kind of information because of *S*'s inability to correctly identify another kind of information does not mean a differential rate of decay for the two kinds of information. The differential, if it exists, is apparent only when *S* must utilize more than one form of information simultaneously. Indeed, there may be three kinds of spatial information to be considered in the Averbach and Coriell paradigm. One is information about the location of a letter, relative to the letter display; another is information about the location of a bar marker, relative to the bar marker display; and a third may be considered as the relative location of the bar display to the letter display. Experiment III demonstrated that *S* had no trouble connecting the spatial location of a bar relative to the bar display if the cue was of a short duration, and thus, in an Averbach and Coriell experiment, the loss of information may also be influenced by *S*'s ability to connect a particular bar location to a particular letter location. In Dick's experiment the apparent loss of spatial information may have resulted from *S*'s inability to connect a letter to its appropriate location in the letter display. An unreported experiment demonstrated that this information does not decay with the delay of the cue.

Information is lost only if *S* has to both locate and identify items in the display. In Experiments II and III, where *S* had to do only one of the two tasks, accuracy was high for all delays of the cue. In Experiment I accuracy was low, with the exception of the short cue delay of 0 ISI. A possible reason for the high level of accuracy of 0 ISI has already been noted. Thus, the decay function found in partial-report experiments may be due not only to a fading icon, but also to the task requirements of the experiment.

REFERENCES

AVERBACH, E., & CORIELL, A. S. Short-term memory in vision. *Bell System Technical Journal*, 1961, **40**, 309–328.

AVERBACH, E., & SPERLING, G. Short-term storage of information in vision. In, *Symposium on information theory*. London: Butterworth, 1961.

DICK, A. O. Visual information processing and the memory trace. Unpublished doctoral dissertation, University of Waterloo, Ontario, Canada, 1967.

DICK, A. O. Relations between the sensory registrar and short-term storage in tachistoscopic recognition. *Journal of Experimental Psychology*, 1969, **82**, 279–284.

ENGEL, G. R. An investigation of visual responses to brief stereoscopic stimuli. *Quarterly Journal of Experimental Psychology*, 1970, **22**, 148–166.

ERIKSEN, C. W. Temporal luminence summation effects in forward and backward masking. *Perception & Psychophysics*, 1966, **1**, 87–92.

ERIKSEN, C. W., & COLLINS, J. F. Some temporal characteristics of visual pattern perception. *Journal of Experimental Psychology*, 1967, **74**, 476–484.

ERIKSEN, C. W., & COLLINS, J. F. Temporal course of selective attention. *Journal of Experimental Psychology*, 1969, **80**, 254–261.

HABER, R. N., & STANDING, L. Direct measures of short-term visual storage. *Quarterly Journal of Experimental Psychology*, 1969, **21**, 43–54.

MACKWORTH, J. F. The relation between the visual image and post-perceptual immediate memory. *Journal of Verbal Learning and Verbal Behavior*, 1963, **2**, 75–85.

MERIKLE, P. M., LOWE, D. G., & COLTHEART, M. Familiarity and method of report as determinants of tachistoscopic performance. *Canadian Journal of Psychology*, 1971, **25**, 167–174.

MEWHORT, D. J. K., & CORNETT, S. Scanning and the familiarity effect in tachistoscopic recognition. *Canadian Journal of Psychology*, 1972, **26**, 181–189.

MEWHORT, D. J. K., MERIKLE, P. M., & BRYDEN, M. P. On the transfer from iconic to short-term memory. *Journal of Experimental Psychology*, 1969, **81**, 89–94.

NEISSER, U. *Cognitive psychology*. New York: Appleton-Century-Crofts, 1967.

SMITH, F., & CAREY, P. Temporal factors in visual information processing. *Canadian Journal of Psychology*, 1966, **20**, 337–342.

SPERLING, G. The information available in brief visual presentations. *Psychological Monographs*, 1960, **74** (11, Whole No. 498).

SPERLING, G. A model for visual memory tasks. *Human Factors*, 1963, **5**, 19–31.

SPERLING, G. Successive approximations to a model for short-term memory. *Acta Psychologica*, 1967, **27**, 285–292.

(Received June 12, 1972)

Journal of Experimental Psychology
1973, Vol. 98, No. 1, 119–124

INTRALIST ORGANIZATION AND SUBSEQUENT FREE RECALLS[1]

OVID J. L. TZENG[2] AND DAISY L. HUNG

Ohio State University

*S*s were trained to overtly rehearse, either forward or backward, every three consecutive input words during list presentation and later asked to free recall the whole list. It was found that these two intralist organization procedures functioned in the same way. That is, relative to the performance of the ordinary free recall *S*s, they had a beneficial effect only on the last three serial positions, but had a detrimental effect elsewhere when the immediate free recall was considered. Furthermore, in the delayed recall, they had a detrimental effect over every serial position. The role of grouping in free recall was discussed from an organizational viewpoint. The conclusion was that "free" rehearsal during acquisition was the basic principle for organization in "free" recall.

The concept of organization plays an important role in contemporary analyses of the free-recall (FR) experiment (Mandler, 1967; Tulving, 1962). There are two basic assumptions underlying organization theory in dealing with the problem of free recall: (*a*) For any list of words presented to a particular *S*, there is an optimal organization that will maximize his recall. (*b*) The organization imposed by *S* on a set of words depends on his perception of the structure of the list (see Postman, 1972). The first assumption refers to the fact that there are numerous strategies which can be employed to organize the list items and that one of the strategies will generate an optimal organization which, in turn, will maximize recall. The second assumption points out the fact that the best strategy demands the *S*'s exposure to the whole list in order to discover any semantic relationship (subjective or objective) embedded in the list. Put together, they imply that any restriction on the part of the *S* in organizing the *whole* list will have a detrimental effect on his recall. This statement is the point of departure for the experiments investigating the negative part–whole transfer in the multitrial free-recall learning situation (Tulving, 1966).

The above statement, according to the very heart of an organization theory, should also be applicable in a single-trial free-recall paradigm. Such a proposition has not yet been too widely investigated. The present study attempted to test one implication of this hypothesis by forcing a particular rehearsal scheme on some of the *S*s. Suppose the *S*s are trained to organize every three consecutive input words during the presentation of the list and, later, asked to free recall the whole list where the performance presumably depends on an optimal organization over the whole list; then, a recall decrement should be expected. For example, a 12-word list is presented to *S* at a 1-sec. rate, and *S* is trained to study each word and, after the presentation of every three words, to overtly recall the preceding three words according to their presented order. At the end of the list presentation, *S* is asked to free recall all of the 12 words in the preceding list. This trained-rehearsal strategy will divide the whole list into four separate blocks (three words per block). If it is true that recall level depends on the integrity of the structure over the whole list, then the restriction imposed by the above intralist-organization procedure will lead to a recall decrement, in comparison to the recall level of the ordinary FR procedure where the integrity of the whole list is presumably possible. By using such an experimental procedure, a recent study by Tzeng, Snyder, and Hung (1972) reported an overall debilitating effect of the

---

[1] The authors are very grateful to Charles N. Cofer for his encouragement during the course of this study. The assistance of Penelope Snavely in correcting some grammatical errors of the writing is acknowledged with appreciation.

[2] Requests for reprints should be sent to Ovid Tzeng, Department of Psychology, Ohio State University, Columbus, Ohio 43210.

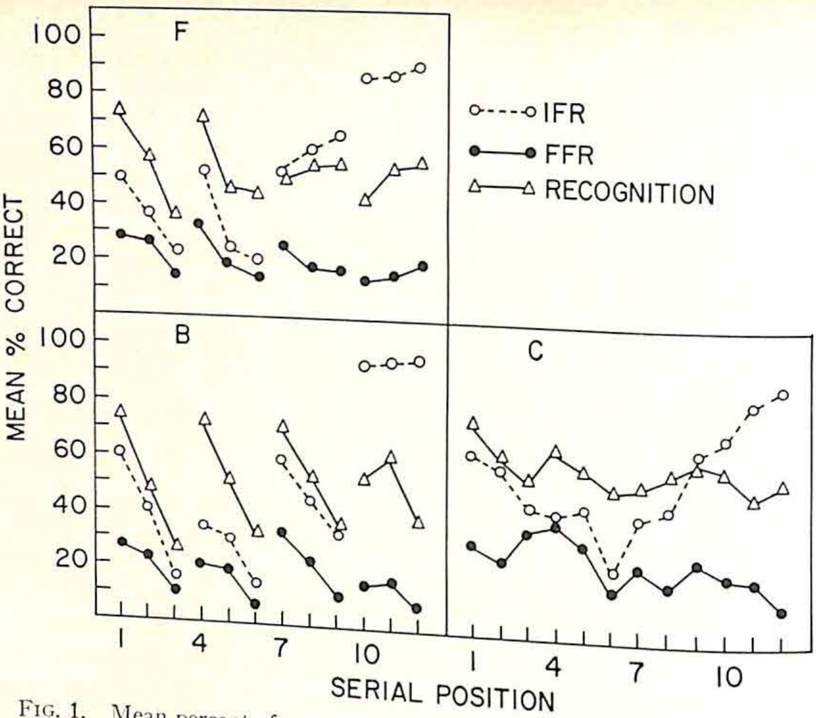


FIG. 1. Mean percent of correct recall on IFR, FFR, and recognition as a function of serial positions for B, F, and C conditions.

intralist organization. However, in that study such an effect seemed to be restricted to the first three blocks (i.e., from Serial Position 1–9). A large facilitating effect was observed at the recency portion (i.e., the last three positions) of the forced-organization group. Since a special FR (i.e., silent rehearsal procedure) was used as the control condition, one might expect the recency effect to be suppressed when the verbalization was minimized. Therefore, the present study employed a more appropriate control condition in order to examine the effect of the intralist organization at the last three positions.

Another interesting finding of Tzeng et al.'s (1972) study was that in the data of the forced-organization group, the recall probability in the first two blocks seemed to be affected by the order of intralist retrieval. That is, the major parameter was the presentation order within each block. It could be argued that this occurred because the *S*s were trained to organize every three words according to their input order. It would be interesting to see if the same response patterns appeared again when *S*s were trained to organize the input words of each block in a reverse order. For example, if the input sequence is A, B, C, then *S* would be asked to reverse the order and call out C, B, A in his intralist organization procedure.

METHOD

Materials. The 120 nouns constituting the experimental word pool were selected randomly from among Thorndike-Lorge words whose frequency of occurrence was between 50 and 100 times per million words. Five unique 12-word lists (sampling without replacement from the pool) were generated for each *S* by an APL/360 computer system.

Subjects. Forty-five undergraduate students enrolled in an introductory educational psychology course at the Pennsylvania State University served as *S*s. They were given additional credits for participation in the experiment.

Procedure. The *S*s were tested individually. They were alternately assigned to one of the three conditions according to the order of their appearance at the laboratory. For the *S* in the *forward intralist organization* (F) condition, the list was presented at a 1-sec rate. He was trained to study the presented words and, after the presentation of every three words, to call out the preceding three words according to their presentation order at a

1-sec. rate paced by an audible "click" (monitored from a Lafayette interval timer). The *S* in the *backward intralist organization* (B) condition was trained to organize every three words in a reverse order. For instance, if the input sequence was A, B, C, then he was to call out C, B, A at a 1-sec. rate. The third condition was an ordinary free-recall condition and served mainly as a *control* (C) condition. The *S* read out every word while the list was presented at a 2-sec. rate in order to equate the total presentation time with the previous two conditions. For all of the above three conditions, an immediate free recall (IFR) was initiated after the complete presentation of the 12 words. Upon the completion of the fifth IFR, *S*s were allowed 3 min. for a final free recall (FFR). Following the FFR, all 120 words in the pool were printed randomly on an 8 × 11 in. sheet, and *S*s were instructed to circle the words they had seen in the five lists.

RESULTS

The overall results, including IFR, FFR, and recognition, for the F, B, and C conditions, are presented in Figure 1. The analysis will be done separately for IFR, FFR, and recognition.

Analysis of IFR Although the performance on the first list was slightly better than that on subsequent lists, analyses of variance (lists as repeated measures) showed no reliable list effect, all $F\text{s} < 1$. Since the particular concern here was to examine the effect of intralist organization on the recall of list words with respect to their serial positions, the mean percentages of correct recalls for the C, B, and F conditions as a function of the four blocks in IFR (summed over the five lists) are summarized in Table 1. Analysis of variance was applied to this 3 Conditions (between-*S*) × 4 Blocks (within-*S*) factorial design. The main effect of blocks was highly significant, $F(3, 126) = 145.28$, $p < .001$. The recency effect was observed in all three conditions. The main effect of conditions was not significant, $F(2, 42) = 1.48$, *ns*. However, the interaction between conditions and blocks, as expected, was statistically significant, $F(6, 126) = 4.16$, $p < .001$. That is, the intralist organization had differential effects over the four blocks. The Scheffé multiple comparison method (Kirk, 1968) was employed to examine this interaction within each block. The contrast $\psi = \mu_{C_i} - \frac{1}{2}\mu_{F_i} - \frac{1}{2}\mu_{B_i}$ (where i referred to the i^{th} block) was tested since the F and B conditions were comparable to each other. For the four blocks, the F values for the contrasts were 10.20, 6.24, .03, and 7.03, respectively. Except for the third block, every F value was greater than the critical value $F'(2,42) = 6.14$, all $p\text{s} < .05$. It was confirmed that the intralist organization had a facilitating effect only on the last block, but had a debilitating effect on the first two blocks.

TABLE 1

MEAN PROPORTION OF CORRECT RECALLS IN IMMEDIATE FREE RECALL FOR THREE CONDITIONS AS A FUNCTION OF FOUR BLOCKS

Condition	Number of block				Total
	1	2	3	4	
C	.57	.40	.55	.83	.59
B	.37	.33	.59	.94	.53
F	.40	.28	.48	.90	.55

Analysis of FFR. Inspection of Figure 1 showed that the FFR of the C condition was better than those of the other two conditions across every serial position. Unlike IFR, the result of FFR seemed to reveal a large list effect. Therefore, the data summarized in Table 2 were collapsed over the four blocks within each list. An analysis of variance for this 3 Conditions (between-*S*) × 5 Lists (within-*S*) factorial design was conducted. Both main effects of conditions and lists were significant, $F(2, 42) = 5.81$, and, $F(4, 168) = 27.87$, respectively, both $p\text{s} < .001$. No interaction between condition and list was observed, $F(8, 168) = 1.17$, *ns*. On the one hand, the consistent superiority of the FFR for

TABLE 2

MEAN NUMBER OF CORRECT RECALLS IN FINAL FREE RECALL FOR THREE CONDITIONS AS A FUNCTION OF FIVE LISTS

Condition	Number of list					Total
	1	2	3	4	5	
C	2.80	2.47	2.33	4.07	5.80	17.47
B	1.20	1.47	2.13	2.67	3.47	10.94
F	1.53	2.27	1.73	2.67	5.07	13.27

TABLE 3

MEAN NUMBER OF CORRECT RECOGNITIONS FOR THREE CONDITIONS AS A FUNCTION OF FIVE LISTS

Condition	Number of list					Total
	1	2	3	4	5	
C	7.20	6.93	6.47	7.73	9.00	37.33
B	7.13	5.40	5.73	6.07	7.06	31.39
F	6.67	6.13	5.80	6.20	7.53	32.33

the C condition over the five lists confirmed the hypothesis that the intralist organization during the input phase had a detrimental effect on the delayed recall where the performance depends, presumably, on the secondary organization (Tulving, 1968). On the other hand, the significant list effect evidenced the retroactive inhibition over lists for unrelated words (cf. Tulving & Psotka, 1971).

Analysis of Recognition. Table 3 summarizes the mean correct recognition scores for the three conditions as a function of the five lists. The data were not adjusted since the averaged numbers of false alarms for the C, B, and F conditions were very comparable to one another. They were 3.8, 4.0, and 3.9, respectively. An analysis of variance similar to the one for the FFR data was performed on those unadjusted recognition scores. Although *S*s in the C condition recognized more words than those of the two organization conditions, the differences were not large enough to yield a significant main effect of conditions, $F(2, 42) = 2.36$. The Conditions × Lists interaction, $F(8, 168) < 1$, was not significant. The only significant effect observed was the list effect. Unlike the results of FFR, both primacy and recency effects appeared in Table 3. Apparently, when recognition was measured, the performance of the earlier lists was released from the strong retroactive inhibition observed in FFR. According to the organization theory, this result indicated that the distinction between availability and accessibility of stored information should be a central issue in the analysis of forgetting (cf. Tulving & Pearlstone, 1966).

Comparison of Response Patterns between B and F conditions. Regardless of the procedural differences of intralist organization, these two conditions, i.e., B and F, produced almost identical results in terms of the overall recall level across the three different memory measures, namely, IFR, FFR, and recognition. Of particular concern here were the response patterns within each of the four blocks. Examination of Figure 1 revealed an interesting result. For the first two blocks, the major response parameter was the presentation order of words within each block. This was true for F condition where the *S*s were trained to organize the words according to their input order. It was also true for B condition where *S*s were trained to organize the words in a reverse fashion according to their input order. It seemed that the function of intralist organization, whether forward or backward, was mainly to have provided *S*s a basis for grouping which functionally separated the items in a list into four blocks so that *S*s were able to devote all their rehearsal activity to a given block. At the time of IFR, this grouping provided effective cues for the words in the later blocks and increased their probability of being recalled. The homogeneous response patterns across the three curves for the first two blocks suggest that the retrieval cues used to generate these items are very stable. On the other hand, the heterogeneous response patterns for the last two blocks suggest that differential retrieval cues are responsible for IFR, FFR, and, maybe, recognition. Hence the results seem to support a cue-dependent interpretation of forgetting (see Tzeng et al., 1972).

DISCUSSION

It has been demonstrated that in FR, during the presentation of a given item, previously presented words are likely to be rehearsed, and the composition of such rehearsal sets varies from one input position to the other (Rundus & Atkinson, 1970). This rehearsal strategy reflects *S*'s intrinsic attempt to organize the input information such that later recall will be maximized. In this sense, providing the freedom for variations in the rehearsal sets during the input phase is the basic principle of organization. Any restriction on that opportunity should cause a recall decrement. The results of the present study on intralist organization confirm this expectation. The recall decrement resulting from forced repetition of a given item at the time of presentation (Glanzer & Meinzer, 1967) can be attributed to the violation of this freedom principle.

The result that the intralist organization plays a debilitating role seems contrary to the grouping effect noted by Neisser (1967). However, several experiments conducted by Gianutsos (1970) consistently demonstrate that in FR grouping has a beneficial effect solely on the last one or two groups. This finding is totally in agreement with an organization interpretation for the FR data. According to Tulving (1968), recency and primacy effects observed in FR represent two types of organization, rather than two different stores (Glanzer & Cunitz, 1966). The first, termed primary, refers to effects attributable to task characteristics; the other, termed secondary, refers to organization that is imposed upon the list by the individual *S*. Accordingly, an organizational theory will predict an interaction between these two types of organization and the grouping effect. That is, in FR, the intralist organization will have a facilitating effect only at the recency portion, but will have a debilitating effect elsewhere. Furthermore, in a delayed test where the recall depends mainly on the secondary organization, intralist organization will have an overall detrimental effect.

Among the many investigators who prefer to talk about the serial position curve as representing two different stores, i.e., short-term memory (STM) and long-term memory (LTM), Glanzer (1972) defines STM as "a very robust, insensitive mechanism that responds to nothing except the passage of items through it [p. 148]." The facilitating effect of intralist organization on the recency portion in IFR is, of course, contrary to this definition. In FFR, not only does this facilitating effect disappear, but performance is actually worse than that of the C condition. These results are not easily accommodated by a two-store hypothesis, which assumes that the recency effect reflects loss from a short-term buffer. If the major parameter in STM is "time" per se, there is no reason to expect that the recall level of organization conditions (i.e., B and F), relative to that of the C condition, changes from extremely good to extremely bad (i.e., from IFR to FFR) while the time factor is held constant for all conditions. Another compensatory process, whatever it is, must be postulated. However, these problematic findings are the necessary expectations for an organization theory of FR. Since the auditory and/or temporal information produced in the intralist organization is readily accessible for IFR, the primary organization will be benefited. On the other hand, intralist organization violates the principle of free rehearsal; thus, the secondary organization will be hindered. Once a certain kind of cue had been generated and used in IFR, the recall probability of FFR would come to depend on the extent to which the *S*s still maintained these cues. The response patterns in Figure 1 support this formulation (see Tzeng et al., 1972).

An organization theory of FR will also argue against the proposition that a nonzero asymptote reflects the "permanent" characteristics of long-term memory. If the asymptote is in any sense permanent, then the stability should hold up across lists (i.e., between trials); otherwise, the stability would hold up only within the framework of a single presentation and test (i.e., within trials). The recency effect substantiated in Table 2 reveals that the asymptote is a "within-trials," rather than a "between-trials" phenomenon. In other words, long-term memory does not seem to be all that permanent (Murdock, 1972). Furthermore, the primacy effect revealed in the recognition data (Table 3) indicates that the above retroactive inhibition is a "cue-depend," rather than a "trace-depend" phenomenon (Tulving & Psotka, 1971).

The conceptualization of the variations in rehearsal composition sets as the basic principle of organization has at least two theoretical implications. First, the hypothesis of "independent trace storage" proposed by Slamecka (1968) becomes enigmatic unless one makes a rigorous assumption that subjective rehearsal during acquisition has nothing to do with "functional storage." This is impractical for it delimits the notion of storage to only the perceptual level. Second, to emphasize the importance of free organization in FR implies that retrieval is not a matter of simple associative chaining, but rather an elaborated plan whose formation demands the breakdown of the presentation order. If this "subjective retrieval plan" is important for optimal recall, then the ineffectiveness (sometimes an inhibitory effect) of contextual cuing observed by Slamecka (1968, 1969), can be interpreted as the degradation of the retrieval plan by the presentation of a random selection of items in a haphazard sequence (Postman, 1972).

In the present study, the experimental paradigm is patterned after Craik's (1970) study wherein *S*s were given several lists to recall one at a time, each such presentation and test being one "cycle." After five such cycles, the *S*s are asked for final recall, i.e., they are to recall as many words as possible from the entire final lists. In a sense, each individual cycle corresponds to a forced intralist organization. Thus, according to the preceding argument, one should expect an impaired FFR relative to the recall level of the *S*s, who simply learn the 60 words presented as a single list. Another 15 *S*s from the same source as previously described were asked to learn the 60-word list presented at a 3-sec. rate. A 30-sec. neutral task was inserted before recall was initiated in order to eliminate the recency effect. The same recognition test followed the recall test. The mean correct recall in this 60-word list was 22.73 which was significantly greater than the FFR of any of the three conditions, $F(3,\ 56) = 16.14$, $p < .001$. The mean correct recognition of the 60-word list (40.07) was also significantly better than that of any of the three conditions, $F(3, 56) = 4.29$, $p < .05$. Therefore, the detrimental effect of intralist organization seems to be very general; and more importantly, this effect is observed not only in recall but also in recognition.

In summary, the main conclusion to be drawn in the present study is that "free" rehearsal in acquisition is the basic principle for organization in "free" recall. Any violation of this principle will impair the long-term retention.

REFERENCES

Craik, F. I. M. The fate of primary memory items in free recall. *Journal of Verbal Learning and Verbal Behavior*, 1970, **9**, 143–148.

Gianutsos, S. Free recall of grouped words. Unpublished doctoral dissertation, New York University, 1970.

Glanzer, M. Storage mechanisms in recall. In G. H. Bower (Ed.), *The psychology of learning and motivation.* Vol. V. New York: Academic Press, 1972.

Glanzer, M., & Cunitz, A. R. Two storage mechanisms in free recall. *Journal of Verbal Learning and Verbal Behavior*, 1966, **5**, 351–360.

Glanzer, M. & Meinzer, A. The effect of intralist activity on free recall. *Journal of Verbal Learning and Verbal Behavior*, 1967, **6**, 928–935.

Kirk, R. E. *Experimental design: Procedures for the behavioral science.* Belmont, Calif.: Brooks Cole, 1968.

Mandler, G. Organization and memory. In K. W. Spence & J. T. Spence (Eds.), *The psychology of learning and motivation.* Vol. 1. New York: Academic Press, 1967.

Murdock, B. B., Jr. Short-term memory. In G. H. Bower (Ed.), *The psychology of learning and motivation.* Vol. V. New York: Academic Press, 1972.

Murdock, B. B., Jr., & Walker, K. D. Modality effects in free recall. *Journal of Verbal Learning and Verbal Behavior*, 1969, **8**, 665–676.

Neisser, U. *Cognitive psychology.* New York: Appleton-Century-Crofts, 1967.

Postman, L. A pragmatic view of organizational theory. In E. Tulving & W. Donaldson (Eds.), *Organization of memory.* New York: Academic Press, 1972.

Rundus, D., & Atkinson, R. C. Rehearsal processes in free recall: A procedure for direct observation. *Journal of Verbal Learning and Verbal Behavior*, 1970, **9**, 99–105.

Slamecka, N. J. An examination of trace storage in free recall. *Journal of Experimental Psychology*, 1968, **76**, 504–513.

Slamecka, N. J. Testing for associative storage in multitrial free recall. *Journal of Experimental Psychology*, 1969, **8**, 557–560.

Tulving, E. Subjective organization in free recall of "unrelated words." *Psychological Review*, 1962, **69**, 344–354.

Tulving, E. Subjective organization and effects of repetition in multitrial free recall learning. *Journal of Verbal Learning and Verbal Behavior*, 1966, **5**, 193–197.

Tulving, E. Theoretical issues in free recall. In T. R. Dixon & D. L. Horton (Eds.), *Verbal behavior and general behavior theory.* Englewood Cliffs, N. J.: Prentice-Hall, 1968.

Tulving, E., & Pearlstone, Z. Availability versus accessibility of information in memory for words. *Journal of Verbal Learning and Verbal Behavior*, 1966, **5**, 381–391.

Tulving, E., & Psotka, J. Retroactive inhibition in free recall: Inaccessibility of information available in the memory store. *Journal of Experimental Psychology*, 1971, **87**, 1–8.

Tzeng, O. J. L., Snyder, H. I., & Hung, D. L. The effect of overt intralist retrieval on subsequent free recalls. *Psychonomic Science*, 1972, **28**, 103–105.

(Received June 12, 1972)

Journal of Experimental Psychology
1973, Vol. 98, No. 1, 125–130

EFFECT OF EXTERNAL TARGET PRESENCE ON VISUAL ADAPTATION WITH ACTIVE AND PASSIVE MOVEMENT[1]

LAWRENCE E. MELAMED,[2] MICHAEL HALAY, AND JOSEPH W. GILDOW

Kent State University

Visual adaptation, produced when *S* observes his prismatically displayed hand moving in front of a homogeneous background, was interpreted as enabling *S* to detect the discrepancy between the visual and proprioceptive localization of the hand. Detection of this discrepancy can be facilitated for passive-movement *S*s by the use of a structured prism-exposure background consisting of line targets serving as external referents for arm localization. A homogeneous background during prism exposure produced significant adaptation only for active-movement *S*s. Both passive- and active-movement *S*s adapted significantly when target lines were included in the exposure background. Adaptation was not dependent upon instructing *S* to specifically track the coincidence between her pointing finger and the target line.

In the early studies by Held and his coworkers (Held & Gottlieb, 1958; Held & Hein, 1958; Held & Schlank, 1959) comparing the efficacy of viewed active and passive movement in adaptation to an optically displaced visual array, the passive-movement conditions did not produce significant adaptation. In these studies, *S* performed both a preexposure and postexposure test of target-pointing accuracy under normal vision. The degree of adaptation achieved by the intervening exposure to an optically displaced visual field was defined by the degree to which the postexposure measurements, relative to the preexposure, were biased in the direction opposite to that of the optical displacement ("negative aftereffect"). In the exposure condition, *S* merely viewed his hand as he moved it or had it moved for him. The hand was viewed on a homogeneous background. The results of these experiments have been explained by a model of the adaptation process in which adaptation involves a recorrelation of the response-produced ("reafferent") sensory information and the efferent signals controlling the movement.

Another view of the activity of *S* in the experiments described above is considered in the present study. Consider that *S* has a well-developed sense of his egocentric orientation when he begins the exposure task. The *S* in the active-movement condition then detects the lateral displacement of the optical array for one of a number of possible reasons. At the beginning of the exposure condition, he is aware of the location of his arm vis-à-vis his body plane. This localization is effected through proprioceptive information. It will differ from the localization of the arm produced by visual input. The *S* will detect this discrepancy between the proprioceptive and visual localizations of his hand if he tracks his movements. Tracking is meant here to be synonymous with testing. The *S* actively tests the location of his hand using the relevant sensory inputs. Thus, he may move his hand such that it appears proprioceptively across his midline, yet visually it appears much further to his left when wearing base-right prisms. Similarly, if *S*'s movements are sufficiently extensive, he may visually detect his hand entering his visual field on his right, but proprioceptively feel it to be much beyond this point to his right. The discrepancy between the apparent visual and proprioceptive locations of *S*'s hand may arise through any number of tracking procedures. Alternatively, passive *S*s may not have been tracking their movements even though this was possible. They were given a very passive "set" to relax and let *E* move their hand for them. They may have simply failed to process the information relevant to the discrepancy between the apparent visual and apparent proprioceptive location of their limb.[3]

---

[1] This research was supported by Grant MH-20504 from the National Institute of Mental Health. The authors wish to thank Terry J. Spencer for many helpful comments on an earlier version of this manuscript.

[2] Requests for reprints should be sent to Lawrence E. Melamed, Department of Psychology, Kent State University, Kent, Ohio 44242.

The focus of this view of visual adaptation is on the necessity of detecting the discrepancy in the localization information about the arm that is being provided by the two sensory channels employed in this task. It may be that the importance of active movement in visual adaptation has been to necessitate that *S* track or test, repeatedly, the sensory inputs from the task that refer to arm location. The fact that the degree of adaptation in the procedure of Held and his co-workers seems to peak in approximately 15 min. (Efstathiou, 1969) probably indicates the limited nature of the information provided in this task. An *S* walking around outdoors for several hours experiences many situations wherethe conflict in visual and proprioceptive information (e.g., stumbling on a curb) is strong and novel, and thus adapts to a much greater degree (Hay & Pick, 1966).

The present study was basically a replication of the Held (Held & Gottlieb, 1958) paradigm in which *S*s observe the active or passive movement of their optically displaced hand. The primary change in procedure was the manipulation of the background upon which *S* observes his hand movements. It was either a homogeneous surface, as has typically been used in this paradigm, or one structured with thin vertical lines. It was predicted that only active-movement *S*s would show significant aftereffects in the homogeneous background condition, whereas *both* active- and passive-movement conditions would produce significant aftereffects when the structured background for movements was provided. These predictions were based on the viewpoint that adaptation in the Held paradigm is derived from the processing of the discrepant visual and proprioceptive information about arm localization. It is assumed that only active-movement *S*s will test for arm position and thus process this discrepant information when the homogeneous exposure background is used. The purpose of the structured exposure background was to facilitate *S*'s ability to test or track the position of his arm. The background supplies reference points for locating the arm with respect to *S*'s own egocentric orientation. It was therefore expected that both passive- and active-movement *S*s would process sufficient information about arm location to show significant aftereffects.

In the structured exposure background conditions, *S*s were required to respond to the visual coincidence of their target finger with the background lines by calling out a number printed underneath the line. In a control condition, passive movement and the structured background were used, but the lines were not numbered and *S*s were not required to report the position of their target finger with reference to the lines. To the extent that passive-movement *S*s using the structured exposure background do adapt, the purpose of the response task and the control condition was to determine if it was necessary to specifically direct the *S*s toward using the target lines in the visual field as localization referents. An active-movement control was not used because it would be expected that any active-movement group in this paradigm would adapt. No specific predictions are offered on the relative size of the aftereffects for the groups expected to produce significant adaptation. This would depend on the specific efficiency of active movement and/or the structured exposure background in providing the necessary sensory inputs.

METHOD

Subjects. The *S*s were 40 female undergraduates recruited from introductory psychology sections at Kent State University. They received course credit

---

[3] It is important to point out that this approach to visual adaptation for the paradigm discussed conflicts with the evidence for a "visual capture" (Hay, Pick, & Ikeda, 1965) of proprioceptive localization. However, the Hay et al. study had *S*s report the apparent location of a static hand as opposed to a constantly moving, tracked hand. The essential element of the studies reviewed was the specific dependence of the negative aftereffect on active movement during the prism exposure period.

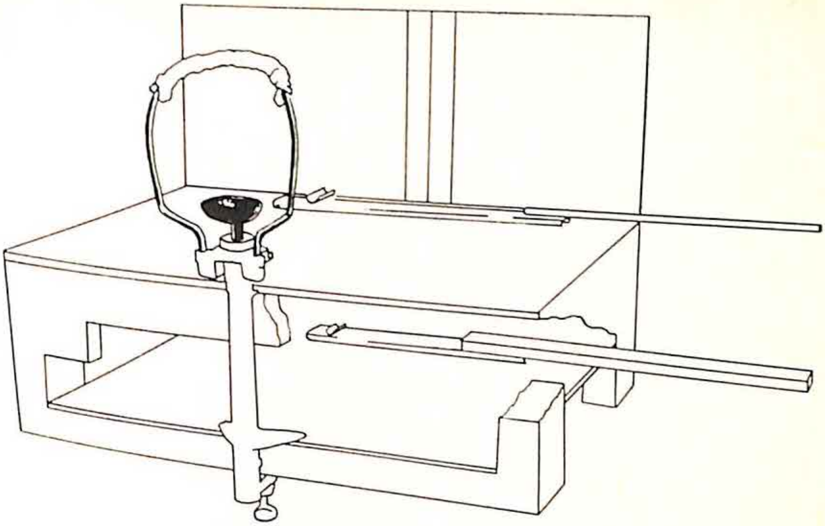


FIG. 1. Test and exposure apparatus.

for participating. Only *S*s who did not wear corrective glasses were selected.

Apparatus. The *S* was seated at a small wooden table with head securely positioned on a combination head-and-chin rest. Throughout all experimental procedures, *S* wore welder's goggles with Risley rotating prisms attached to the front surface. A wide cloth band was attached to the goggles and the head rest in order to further restrict head movement. The binocular field of view was approximately 45°. In front of *S*, on the table, was a 49.53 × 77.47 cm. rectangular wooden box, open on the side facing *S* (see Figure 1). When *S* was to observe movements of her hand, it was placed in an aluminum holder that ran on an aluminum track on top of the box. Immediately behind the track was a wooden backboard that extended the entire width of the box. The *S*'s hand was visually coincident with the backboard. For the tests involved in measuring the negative aftereffects resulting from prism viewing of hand movement, a white tagboard was placed over the backboard. As shown in Figure 1, three black lines, .32 × 15.24 cm. and 5.08 cm. apart, were centered on the tagboard. During these tests, *S*'s hand was placed in a holder on an aluminum slide in the interior of the box. The slide was directly below the slide on the surface of the box. Given the placement of *S*'s head, her field of view, and her proximity to the position of the box on the table, no part of her arm or hand was visible in the test conditions. The location of *S*'s judgment of a target line was determined by reading a value from a measuring stick that was attached to the interior slide and thus moved with it. In certain cases of the exposure conditions, in which *S* viewed her own hand movement on top of the box, a white tagboard containing eight .32 × 15.24 cm. black lines, 2.54 cm. apart, was attached to the backboard of the apparatus.

Design. The primary experiment featured a 2 × 2 factorial design in which active vs. passive movement during exposure to the lateral displacement produced by the prisms was one factor, and a homogeneous exposure background vs. numbered target lines as the exposure background was the other. Eight *S*s were randomly assigned to each of the four groups. Direction of optical displacement was counterbalanced in each group by having four of the *S*s wear prisms with the bases oriented right and the other four with the bases oriented left. In addition, a control group of eight *S*s used passive movement and an exposure background consisting of unnumbered target lines.

Procedure. The experiment was conducted over a 2-day period. On the first day, the *S*s practiced pointing their unobserved hand at target lines for approximately 20 min. This was done in an attempt to produce a stable base line for the pointing tests used to assess the effect of the prismatic viewing.

On the second day, all *S*s underwent a pointing-accuracy task before and after a 10-min. exposure to a 20-diopter lateral prismatic displacement. The Risley prisms were set at 0 diopters for the pre-exposure and postexposure pointing tasks. In each of these tasks, the *S* pointed five times, in a random order, at each of three vertical black lines centered on a white tagboard sheet placed on the backboard of the apparatus. The *S*'s hand was not visible as it was located on the internal slide.

All of the experimental manipulations took place during the prism-exposure period. In the passive-movement conditions, the *S* watched her hand moved laterally across her visual field by *E* at a rate of about one cycle every 6 sec. The *E* accomplished this using the metal rod attached to the slide on the top surface of the apparatus box. In the active conditions, *S* moved her hand herself at the same pace. An electronic metronome was used to indicate the movement rate. In all exposure conditions, *S*'s hand was situated on top of the rectangular box. The extent of movement was equated over all movement conditions. Total movement was approximately 33.02 cm. across the visual field, divided equally to the right and left of *S*'s midline. In the active-movement conditions, the points of movement were indicated to *S* by *E* using the word "stop" as *S* approached the appropriate location. For passive-movement conditions, *E* used guide markers on the metal rod used in moving *S*'s hand. The extent of movement was such that it proceeded 7.62 cm. beyond the edges of the line targets in the target-line exposure background conditions. In the homogeneous exposure background condition, *S* viewed the movement of her hand against a white background produced by attaching a white tagboard sheet to the backboard. However, in the target-line exposure background condition, this movement was viewed against the background of eight vertical black lines. For the experimental groups, the lines were numbered sequentially, from right to left, on the bottom. The *S* in both the active- and passive-movement conditions was required to call out the number of each line as her finger first contacted it. In a control condition, passive-movement *S*s were instructed to watch their hands as they were moved by *E* in front of the eight lines. However, the lines were not numbered.

RESULTS

The aftereffects of the prism-exposure period constituted the data for analysis. These aftereffects are difference scores that were obtained by taking the average postexposure pointing response for each target for each *S* and subtracting the comparable preexposure measure. For purposes of analysis, aftereffects were considered positive when they were in the direction opposite to the lateral displacement produced by the prisms. (This has traditionally been called a "negative aftereffect.") A mixed design analysis of variance was performed in which the between-*S* factors were exposure movement (active or passive), exposure background (target line or homogeneous), and prism base orientation (left or right base) while the within-*S* term was test target (right, center, or left). The analysis of variance indicated that none of the main effects were significant: exposure movement, $F\,(1, 24) = .01$; exposure background, $F\,(1, 24) = 1.17$; prism base orientation, $F\,(1, 24) = 1.64$; and test target, $F\,(2, 48) = .33$. The interaction of central interest, that of Exposure Movement × Exposure Background, was also not significant, $F\,(1, 24) = 1.38$.

The one interaction that was significant and accounted for a large proportion of the variance in this analysis was Prism Base Orientation × Test Target, $F\,(2, 48) = 8.92$, $p < .001$. The data indicate that aftereffects were enhanced when the test target and the direction of the lateral shift produced by the prisms were on the same side of the visual field. Thus, where base-left prisms were used, which shifted the visual field 11° to the *S*'s *right*, the right, center, and left test targets produced aftereffects of 2.25°, 1.54°, and .25°, respectively. A Newman-Keuls analysis indicated that the right–left difference was significant ($p < .01$) as was the center–left difference ($p < .05$). Conversely, when base-right prisms were worn that shifted the visual field to *S*'s *left*, the right, center, and left test targets produced aftereffects of 1.14°, 2.61°, and 3.36°, respectively. The right–left difference ($p < .01$) and center–right ($p < .05$) differences were significant by the Newman-Keuls test. Given that none of the higher order interactions with Prism Base Orientation × Test Target were significant, this interaction was independent of exposure background and movement factors.

Traditionally, in studies of visual adaptation, the size of aftereffects within the experimental groups themselves are evaluated for their significance from zero. When aftereffects are used as the dependent measure, *t* tests for correlated means are employed to compare the preexposure and postexposure measures for each group. This analysis was applied to the data of the present study for the comparisons in which differential outcomes were predicted, i.e., the groups involved in the Exposure Movement (active or passive) × Exposure Background (homogeneous or target line) interaction. The results of one-tailed *t* tests for the four group means are presented in the right-hand column of Table 1. Since the Prism Base × Test Target interaction described above was evident for all groups and was not a differential finding, it indicates that the aftereffects represented by the group means were primarily due to enhanced aftereffects for the target line

TABLE 1

MEAN AFTEREFFECTS FOR THE THREE TEST TARGETS AS A FUNCTION OF PRISM BASE ORIENTATION FOR THE FOUR EXPERIMENTAL GROUPS

Exposure background	Exposure movement	Prism base orientation	Test target			*M*[a]	*SD*[a]
			Left	Center	Right		
Homogeneous	Active	Right	3.22	2.75	.95	1.92**	1.32
		Left	.75	2.80	1.08		
	Passive	Right	2.70	1.98	1.90	.93	2.36
		Left	−1.72	0.00	.75		
Target line	Active	Right	3.30	3.12	2.00	1.85*	1.73
		Left	.25	.85	1.58		
	Passive	Right	4.22	2.58	−.48	2.70*	2.70
		Left	1.75	2.52	5.60		

[a] Mean and standard deviation collapsed over test target and prism base orientation.
* $p < .025$.
** $p < .005$.

on the same side as the prismatic shift. This interaction is presented in Table 1 for each of the four groups. The results of the *t* tests indicate that with a homogeneous exposure background, the typical findings of the early Held studies (Held & Gottlieb, 1958; Held & Hein, 1958; Held & Schlank, 1959) were replicated. The active-movement group showed an average aftereffect of 1.92°, which was significant, $t\ (7) = 3.68$, $p < .005$. The passive-movement group had an average aftereffect of .93°, which was nonsignificant, $t\ (7) = 1.04$. With the target-line exposure background, both the active- and passive-movement groups showed significant aftereffects. The average effect for the active-movement group, 1.85°, $t\ (7) = 2.83$, $p < .025$, was similar to that obtained with the homogeneous background. The average aftereffect for the passive-movement group was 2.70°, $t\ (7) = 3.65$, $p < .025$. Although not represented in Table 1, the control group, using the unnumbered target lines as an exposure background, showed a sizable aftereffect of 2.11°, $t\ (7) = 2.24$, $p < .05$. The standard deviations (*SD*s) of the aftereffect measures for the four experimental groups are given in Table 1. Although the *SD*s for passive movement are larger than those for active movement, this measure did not relate to the size of the aftereffects. Both the group showing the least adaptation and that showing the most were passive-movement groups.

DISCUSSION

A model of visual adaptation has the task of defining commonalities in the experimental conditions of the present study in which adaptation occurred. These findings are not consistent with Hein and Held's (1962) model of the adaptation process in which reafferent sensory information is necessary to produce adaptation. Only active movement produced significant aftereffects when the homogeneous exposure background was used. However, both active- and passive-movement groups adapted significantly with the target-line exposure background. These results were predicted on the assumption that adaptation in the Held paradigm is derived from the processing of dissimilar visual and proprioceptive information about arm localization. This processing necessitates that *S*s test for arm position, both visually and proprioceptively, during the exposure condition. The purpose of the target-line exposure background was to provide external reference points for tracking the position of the arm, i.e., enhance *S*'s ability to test for position. In the absence of a structured exposure background, e.g., the homogeneous background of the Held paradigm, only active-movement *S*s would be expected to track arm position given that passive-movement *S*s have no need to control and direct movement. The results further indicate that adding a structured background does not enhance adaptation for active-movement *S*s. The question then arises about the manner in which active-movement *S*s test for arm position.[4] The significant Prism Base

---

[4] The analysis of the Prism Base Orientation × Test Target interaction was suggested by Benjamin

Orientation × Test Target interaction may indicate one way in which this is accomplished. The results indicate that significantly greater aftereffects were found, in *all* experimental conditions, for the test target located on the side of the visual field toward which the prismatic shift occurs, in comparison with the target on the opposite side of the visual field. This finding implies that visual boundaries can serve as external "targets" in the adaptation process. The term "visual boundaries" refers to the fact that the side of the visual field to which the prismatic shift is directed becomes truncated. This is probably due to the fact that the eyes shift their direction of gaze toward the prism apex. None of the higher order interactions involving the prism base orientation and test target factors were significant. Thus, the fact that this interaction was not unique to a particular segment of the adapting groups would seem to indicate that when *S*s track movement within the visual field, even when involving passive movement and when induced by the background lines, the visual boundaries become enhanced as points for testing arm location.

It may be appropriate to relate the findings of the present study to those obtained by Singer and Day (1966). In that study, both active- and passive-movement *S*s showed significant aftereffects from prism exposure. The study employed procedures that do not appear easily distinguishable from the methods employed in other studies, such as in the homogeneous exposure background conditions of the present experiment, where passive *S*s failed to adapt. However, there was one difference: Singer and Day's *S*s had two target lines present in the visual field in which they were observing their finger movements. These lines were used in the preexposure and postexposure coordination test, but were still visible during the exposure period. It is interesting to speculate that these lines could have served the same purpose as the target lines in the present study. This argument is based on the sizable aftereffects found for the passive-movement group using the unnumbered target lines. This latter finding indicates that the numbering of the target lines and the attend instructions were probably not as essential in the current research as the presence of the line targets.

REFERENCES

Efstathiou, E. Effects of exposure time and magnitude of prism transform on eye–hand coordination. *Journal of Experimental Psychology*, 1969, **81**, 235–240.

Hay, J. C., & Pick, H. L. Visual and proprioceptive adaptation to optical displacement of the visual stimulus. *Journal of Experimental Psychology*, 1966, **71**, 150–158.

Hay, J. C., Pick, H. L., & Ikeda, K. Visual capture produced by prism spectacles. *Psychonomic Science*, 1965, **2**, 215–216.

Hein, A., & Held, R. A neural model for labile sensorimotor coordinations. In E. E. Bernard & M. R. Kane (Eds.), *Biological prototypes and synthetic systems.* Vol. 1. New York: Plenum Press, 1962.

Held, R., & Gottlieb, N. Technique for studying adaptation to disarranged hand–eye coordination. *Perceptual and Motor Skills*, 1958, **8**, 83–86.

Held, R., & Hein, A. V. Adaptation to disarranged hand–eye coordination contingent upon re-afferent stimulation. *Perceptual and Motor Skills*, 1958, **8**, 87–90.

Held, R., & Schlank, M. Adaptation to disarranged eye–hand coordination in the distance dimension. *American Journal of Psychology*, 1959, **72**, 603–605.

Singer, G. H., & Day, R. H. Spatial adaptation and aftereffect with optically transformed vision: Effects of active and passive responding and the relationship between test and exposure responses. *Journal of Experimental Psychology*, 1966, **71**, 725–731.

Wallace, B. An examination of the role of attention and visual-proprioceptive information on visual adaptation to prismatic displacement. Unpublished doctoral dissertation, Kent State University, 1972.

(Received June 14, 1972)

Wallace. This interaction has been replicated under a variety of exposure conditions (Wallace, 1972).

Journal of Experimental Psychology
1973, Vol. 98, No. 1, 131–137

RELEASE OF PROACTIVE INTERFERENCE AS A RESULT OF CHANGING PRESENTATION MODALITY

NEAL E. A. KROLL,[1] JOAN BEE, AND GABRIELE GURSKI
University of California, Davis

Release from proactive interference was measured following shifts in presentation modality. A large release effect was found both when shifting from visual to auditory memory letters and from auditory to visual memory letters. This release was observable both in terms of (*a*) an increase in the probability of a correct recall response and (*b*) a greater likelihood of intrusions in the control conditions coming from the memory trigram of the previous trial. The present finding of a release effect following modality shifts, while earlier findings were inconsistent, indicates that the use of shadowing as the distractor task is likely to result in memory codes isomorphic with memory stimulus presentation.

Wickens (1970) has developed a technique for measuring how to-be-remembered stimuli are encoded in memory. The technique assumes that if a stimulus is represented by a memory code similar to the code used to represent a memory stimulus presented on the previous trial, proactive interference from this earlier code will reduce the probability of correctly recalling the memory stimulus of the current trial. Additional trials with the same type of stimulus increase the proactive interference, thereby reducing even further the probability of a correct response. Any change in the type of memory stimulus which results in a major change in the type of memory code employed should greatly reduce the amount of proactive interference from earlier trials and thereby increase the probability of a correct response. Thus, the improved performance found after a change in the type of memory stimulus is evidence for a "release of proactive interference" and the trial on which the change occurs is called the "release trial."

Recently, Hopkins, Edwards, and Gavelek (1971) employed this technique in an attempt to establish the importance of stimulus modality on the memory code. While they found a large recovery from proactive interference when they shifted from visual to auditory presentation of the memory stimulus, they found no recovery at all in shifting from auditory to visual. On the basis of their results, Hopkins et al. suggest that it may be necessary either to (*a*) drop the assumption that visual and auditory stimuli are encoded differently, or (*b*) drop the assumption that the Wickens (1970) technique does, in fact, measure the important features of the memory code. However, while it may be true that *S*s occasionally do not encode visual and auditory stimuli differently (e.g., forming subvocal or word-association codes for both) and may not have done so in the Hopkins et al. study, it does not necessarily follow that this is the usual case. There is considerable evidence that *S*s can, given the incentive to do so, use either name or visual image memory codes for *either* verbal or pictorial memory stimuli (e.g., Tversky, 1969). It would seem, therefore, that to ensure a release from proactive interference resulting from a switch in presentation modality one would have to be certain that the experimental conditions were conducive to using memory codes isomorphic with the method of presentation. One procedure which apparently has this property is having *S* shadow (i.e., repeat aloud) a series of spoken letters during the retention interval following the presentation of either visual or auditory memory letters. Under these conditions, *S*s apparently use name codes (i.e., subvocal responses and/or auditory imagery) for the auditory but not the visual memory letters (Parkinson, Parks, & Kroll, 1971; Salzberg, Parks, Kroll, & Parkinson, 1971). However, visual imagery is apparently

---

[1] Requests for reprints should be sent to Neal E. A. Kroll, Department of Psychology, University of California, Davis, California 95616.

TABLE 1

MEMORY TRIGRAM PRESENTATION MODALITY SEQUENCES USED IN EXPERIMENT I

Sequence	Trial								
	1	2	3	4	5	6	7	8	9
I	A	V	V	V	V(C_2)	V	V	V	A(R_3)
II	A	V	V	V	A(R_1)	A	A	A	A(C_3)
III	V	A	A	A	A(C_1)	A	A	A	V(R_4)
IV	V	A	A	A	V(R_2)	V	V	V	V(C_4)

Note. Abbreviations: A = auditory memory trigram presentation; V = visual memory trigram presentation; (R_i) = release trial; and (C_i) = control for (R_i).

used for remembering the visual memory letters (Parks, Kroll, Salzberg, & Parkinson, 1972).

The present experiments measure the release effect of switching presentation modality when *S* is required to shadow during the retention interval. If the shadowing procedure is successful in its attempt to induce *S* to employ isomorphic memory codes for visual and auditory presentations, a strong release effect should be found when the presentation modality is switched in either direction.

EXPERIMENT I

Method

Subjects. Undergraduate volunteers (126 females and 66 males) from an introductory psychology course at the University of California, Davis, served as *S*s for this experiment and received class credit for their participation.

Design. There were 32 groups of six *S*s each. The 32 groups represented all possible combinations of four different sequences of presentation modality (of the memory trigram), two different sequences of shadow lists, and four different sequences of memory trigrams. Table 1 shows the four different sequences of presentation modality. Two sequences (I and III) have the release trial on Trial 9, while the other two sequences (II and IV) have the release trial on Trial 5. Thus, every *S* was observed under both control and release conditions, but on different trials. The first trial of every sequence used the same presentation modality as the release trial of that sequence in order to reduce possible startle effects. There were two different sequences of shadow lists with half of the *S*s in each of the presentation modality sequences shadowing numbers on the first five trials and shadowing letters on the last four trials and the other *S*s shadowing letter lists before number lists. Within each of the resulting eight combinations of presentation modality sequences and shadow list sequences, there were four different sequences of memory trigrams. These four different assignments of memory trigrams to trials were used to help equate trigram difficulty level across trials.

Materials and apparatus. Shadow lists were constructed by randomly selecting 35 letters (from the set A, C, D, I, M, O, R, S, U, and Y) or 35 numbers (1–10) with the restriction that no particular letter or number was used twice in succession. A female reader recorded the resulting lists at the rate of two letters or numbers per second. The memory trigrams were selected from the letters B, F, G, H, J, K, L, N, Q, T, X, and Z with no letter occurring more than once in any particular trigram. The association value of each trigram was less than or equal to 29% (Witmer, 1935). One track of a stereo tape recorder was used to present the shadow lists and the auditory memory letters binaurally via earphones, while the other track contained signals which controlled the visual presentations by means of a voice-sensitive relay. A card punch (IBM Model 026) and associated relay and timing circuitry controlled the sequencing of the visual letters. The visual letters (8.5 cm. in height) appeared via a single one-plane readout unit (Industrial Electronic Engineers, Series 80000) located approximately 1.8 m. from *S* at eye level. The three letters of each trigram were flashed at a rate of 2/sec, with an on time of approximately 200 msec. each.

Procedure. Each *S* reported individually and was assigned to one of the groups according to a prearranged schedule based on the order of arrival. Practice was given on both number and letter shadowing until an 80% accuracy level was demonstrated, after which *S* received instructions for the memory task. On the auditory memory trials, the memory trigrams were recorded in a male voice and replaced three of the female-voiced letters or numbers of the shadow task. Since the ninth, tenth, and eleventh items were replaced, there was a 4-sec. lead-in to the memory trigram and a 12-sec. retention interval. The *S*s were instructed to shadow both the female- and male-voiced letters or numbers and to remember the male-voiced letters. On the visual memory trials, the three letters of each trigram were flashed successively, contiguous with the female voice reading the ninth, tenth, and eleventh shadow letters. The *S*s were instructed to continue shadowing the female voice, while watching and remembering the visually presented letters. At the beginning of each memory trial, a tape-recorded message told *S* whether the memory trigram would be presented visually or auditorily on that particular trial. After writing his response, *S* slid his response sheet under a black cover so that subsequent responses would not be directly influenced by earlier responses. In order to insure that the response period would not change for the last trial and to prevent a change in alertness that might occur if *S* knew he was on the last trial, the response sheet contained spaces for 15 trials and *S* heard the beginning of a tenth trial before the end of the experiment was signaled. Thus, the end of the response period for Trial 9, like those

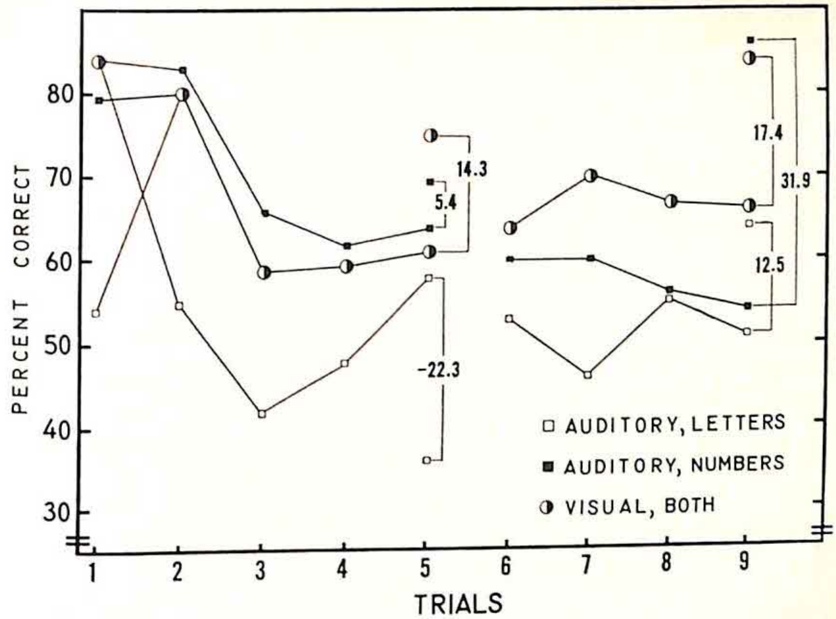


FIG. 1. Percentage of correct letters recalled in Experiment I as a function of memory trigram presentation modality, type of shadow list, and trial number.

of earlier trials, was indicated by the beginning of another trial.

Results and Discussion

The recall responses were scored with a "free-recall" criterion; i.e., order of recall was not scored, only the recall of each individual memory letter. The data were also scored on the basis of serial recall (a letter was counted as correct only if recalled in the proper position), on the basis of perfect recall (a trigram was scored as correct only if all three letters were recalled in correct order), and on a 4-point basis (1 point for each correct letter and an additional point if all three letters were recalled in the correct order). Extremely similar patterns were found by all scoring methods. Figure 1 shows the resulting percentage correct as a function of presentation modality of the memory trigram and trial number. As in an earlier experiment (Kroll, 1972), the type of shadow list (letters vs. numbers) greatly influenced the recall performance with auditory memory letters, but had little or no effect with visual memory letters. Therefore Figure 1 presents recall of auditory memory letters as a function of type of shadow list, but averages across the two types of shadow lists for recall of visual memory letters. Performance is also averaged over all *S*s having the same presentation modality for all trials except Trials 5 and 9 where control and release conditions are presented separately. In Figure 1, the control conditions are connected with the points representing earlier trials having the same presentation modality. However, the release conditions are connected to their appropriate controls with a vertical line labeled by the difference between the percentage correct for the release condition and the percentage correct for the control condition. (Refer to Table 1 for the different sequences of presentation modality.)

Since earlier experiments found markedly different release effects when switching from auditory to visual presentation than when switching from visual to auditory presentation, these conditions were analyzed separately. Trial 5 was also analyzed separately from Trial 9. Thus, four separate 2×2 analyses of variance were performed. In each, the orthogonal factors were type of shadow list (letters vs. numbers) and condition (release vs. control).

On Trial 5, recall of the visual memory letters was significantly better in the release condition, $F\ (1, 92) = 4.870, p < .05$, but was not influenced by type of shadow list ($F < 1$) nor by the Condition × Shadow List interaction ($F < 1$). The pattern for recall of auditory memory letters on Trial 5, however, is much more complicated. As shown in Figure 1, there appears to be a small release effect with number shadowing, but a large "negative release effect" with letter shadowing. That is, *S*s with letter shadowing and auditory memory letters on Trial 5 remembered less if they had just switched from visual memory letters than if they had been receiving auditory memory letters all along. This pattern results in a significant interaction, $F\ (1, 92) = 5.637, p < .025$, and poorer recall with letter shadowing, $F\ (1, 92) = 11.049, p < .005$, but no significant direct effect of condition, $F\ (1, 92) = 2.029, p < .10$. This negative release effect is difficult to understand, but may be largely due to a chance increase in control performance on Trial 5. (Notice that performance of control *S*s with auditory memory letters and letter shadowing is much higher on Trial 5 than on Trials 3 or 4.) Also *S*s in the release condition may have found it difficult to handle the sudden increase in task difficulty (i.e., most *S*s find remembering auditory memory letters much more difficult than remembering visual memory letters while shadowing letters).

On Trial 9, recall of the visual memory letters was again significantly better in the release condition, $F\ (1, 92) = 10.684, p < .005$, while not being influenced by either type of shadow list ($F < 1$) or the interaction ($F < 1$). Recall of auditory memory letters on Trial 9 also shows better recall in the release condition, $F\ (1, 92) = 12.383, p < .005$. Unlike Trial 5, the interaction is not significant, $F\ (1, 92 = 2.370, p < .10$, while the effect of type of shadow list on auditory recall is now of borderline significance, $F\ (1, 92) = 3.928, p \simeq .05$.

On Trial 9 there appears to be 100% recovery (return to Trial 1 performance) in all four of the release conditions (visual and auditory memory letters, number and letter shadowing). The degree of the recovery may be somewhat exaggerated, however, since the memory trigrams used on Trial 9 were not counterbalanced with those used on Trial 1 and the former may have been somewhat easier. While the size of the recovery (in terms of return to Trial 1 performance) is somewhat suspect, the occurrence of the release effect (in terms of difference between release and control performance) is not. The exact same memory trigrams and shadow list recordings were used for both control and release conditions.

If errors increase over trials because of a buildup of proactive interference, then later trials should show a high percentage of intrusions from preceding trials. Conversely, if a shift in the modality of memory letter presentation reduces proactive interference, release trials should show a low percentage of intrusions from preceding trials. That this, in fact, happened in the present experiment is suggested by the finding that 24.3% of the intrusions on Control Trials 5 and 9 were from the memory trigram of the immediately preceding trial (30.5% for visual presentation, 20.2% for auditory presentation), while only 16.9% of the intrusions on Release Trials 5 and 9 were from the preceding trial (17.5% for visual, 16.5% for auditory). The reliability of this change in the source of intrusions was tested by comparing the frequencies, within control and release conditions (Trials 5 and 9), of *S*s giving at least one intrusion from the previous trial (Trial 4 or 8). The chi-squares reported below are based on the more conservative form of these tests where the "failure to give at least one intrusion from the previous trial" column of the 2 × 2 table included only those *S*s who made at least one error on the control or release trial. When control and release trials used visual presentation, the relationship between condition and source of intrusion was significant, $\chi^2\ (1) = 12.922, p < .001$. However, this relationship was not significant with auditory presentation, $\chi^2\ (1) = .834$. This lack of significance was

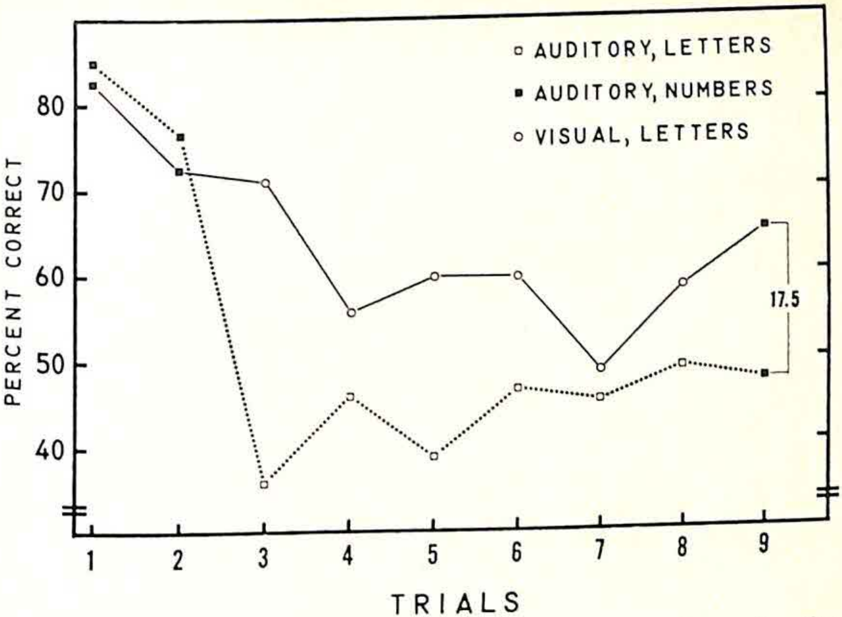


FIG. 2. Percentage of correct letters recalled in Experiment II as a function of memory trigram presentation modality, type of shadow list, and trial number.

largely due to the large negative release effect on Trial 5.

EXPERIMENT II

Experiment I found a strong, consistent release effect when going from auditory to visual memory letters. However the effect following a shift from visual to auditory memory letters was inconsistent. While the negative release effect found on Trial 5 was probably a chance occurrence, it may reflect some underlying difficulties which make the procedure inappropriate for the assessment of the release effect. Therefore it seemed necessary to demonstrate that shifting from visual to auditory memory letters does indeed result in release from proactive interference when shadowing is used as the intervening task. Experiment II, therefore, is essentially an attempt to replicate the release effect found on Trial 9 of Experiment I when shifting from visual to auditory memory letters.

Method

The *Ss* were 29 male and 51 female undergraduate volunteers from University of California, Davis, introductory psychology classes who received class credit for their participation. The *Ss* were assigned to one of two conditions on the basis of their order of arrival. In both the conditions, Trials 1, 2, and 9 had auditory memory letters and number shadowing, while Trials 3–8 in both of the conditions had letter shadowing. The conditions differed only in the method of presenting memory letters on Trials 3–8, with the control condition using auditory presentation and the release condition using visual presentation. To provide an accurate measure of the degree of release recovery, half of the *Ss* in each condition received as their Trials 1 and 9 the same recordings that the other *Ss* received, but in reversed order (i.e., as Trials 9 and 1). In all other respects, the materials, apparatus, and procedure were the same as in Experiment I.

The first two trials used the same type of shadow list and the same presentation modality as used in the last trial (Trial 9) so that *S* would be well familiar with the task (number shadowing and auditory memory). Trial 9 was used as the release trial to ensure maximum dissipation of proactive interference from Trials 1 and 2. Number shadowing was used on Trial 9 since letter shadowing apparently causes a great deal of retroactive interference for auditory memory letters in addition to reducing the number of rehearsals possible (Kroll, 1972). This additional source of difficulty may reduce the amount of release possible. Letter shadowing was used on Trials 3–8 to provide greater protection from *Ss* recoding the visual memory stimuli into subvocal memory codes.

Results and Discussion

The results are presented in Figure 2. As on Trial 9 in Experiment I, there was a significant release effect, t (78) = 2.865, $p < .01$. The magnitude of this effect was increased slightly if difference scores (Trial 1 correct − Trials 9 correct) were used instead of raw scores, t (78) = 2.928, $p < .005$. The degree of recovery, however, is much smaller in Experiment II (54.3%) than in Trial 9 of Experiment I (approximately 100%). This difference between the degree of recovery found in Experiments I and II is probably due to the counterbalancing technique used in Experiment II and the lack of it in Experiment I.

The release effect was also evident in the Trial 9 intrusion data with 23.8% of the intrusions in the control condition coming from the three letters used as memory letters on the previous trial. Of those *S*s giving at least one intrusion, more *S*s in the control condition gave intrusions from the previous trial than did *S*s in the release condition, χ^2 (1) = 4.442, $p < .05$.

General Discussion

The results from the two reported experiments support Wickens' (1970) contention that a switch in *S*'s method of encoding the memory stimulus results in a release of proactive interference. The present experiments affected this encoding by changing the presentation modality under conditions where *S* is likely to encode the stimulus in a way analogous to the modality of presentation. Earlier experiments which failed to find a release following changes in presentation modality may have been using procedures less likely to ensure such encoding (e.g., using a less vivid visual stimulus or using a distractor task that places a heavy burden on processes involved with visual imagery). That the observed change in percentage correct is due to a release phenomenon, rather than a startle or isolation effect (e.g., Gumenik & Rossman, 1970), is supported by the dramatic change in the source of intrusions.

The present experiments also changed the distractor task from letter shadowing to number shadowing and vice versa with no apparent release effect resulting from this change (see Trial 6, Figure 1). This lack of a release effect following a change in distractor task is additional (although perhaps weak) evidence that the release effect cannot be attributed to an increase in attention, since a change in distractor task should produce at least some novelty and increased attention. A release effect would also be expected if a change in the distractor task affected *S*s method of encoding the memory stimulus. Thus, while these two distractor tasks differentially affect the *efficiency* of encoding the memory stimulus, they apparently do not differentially affect the *method* of encoding.

It is interesting to note that proactive interference builds up approximately to the same extent with both distractor tasks. (If this were not true, then the Trial 9 performance of the control group in Experiment II should have been nearly as good as on Trials 1 and 2.) This is somewhat surprising given that the probability of correct recall of auditory memory letters is much higher with number shadowing than with letter shadowing. Thus, while letter shadowing reduces the probability of a correct response, it still allows the buildup of proactive interference. One explanation of this would be that *S*s confuse their subvocal rehearsals of the auditory memory letters with the shadow letters (Kroll, 1972; Kroll & Kellicutt, 1972), thereby often remembering the incorrect letters, and that these incorrect memories produce as much proactive interference as do correct ones. A closer examination of the intrusions made by the control group on Trial 9 of Experiment II supports this explanation: 81.2% of those intrusions that were Trial 8 memory letters had been remembered by *S*s, and 35.4% of those intrusions that were not Trial 8 memory letters had been given as intrusions on Trial 8. (Comparable figures for the release group were 75.0% and 14.5%.)

The following model, then, is proposed for explaining how *S*s remember and why they forget the memory letters in experiments using shadowing as the distractor task. When the memory letters are presented visually, *S*s tend to encode them visually (Parks et al., 1972) and are able to rehearse this visual image during the retention interval relatively efficiently (Kroll & Kellicutt, 1972). An increase in difficulty level of the shadow material reduces the probability of a correct response (Parkinson et al., 1971), probably by reducing the number of rehearsals possible. However, increasing either name or class similarity between memory letters and shadow material does not affect the probability of a correct response (Kroll, 1972; Parkinson et al., 1971; Salzberg et al., 1971), as it should not, given the great qualitative difference between the visual memory code and the auditory and vocal nature of the shadow task. However, the present experiments suggest that memory for visual memory letters is reduced by proactive interference from earlier visual memory trials. This may be caused by *S*'s confusing his memory code for the current trial with his memory code for an earlier trial. If this is true, then one of the major functions of rehearsal in this type of task is probably to differentiate the present memory code from earlier memory codes, perhaps by adding a "recency" tag.

When the memory letters are presented auditorily, *S*s tend to encode them auditorily or subvocally. As shown in these experiments, the auditory memory codes are subject to proactive interference from earlier auditory trials, much in the same manner as are visual codes subject to proactive interference from earlier visual trials. However, since the auditory codes are similar in form to the ongoing shadow task, auditory rehearsals are less efficient than are visual rehearsals (Kroll & Kellicutt, 1972), since they are more likely to be confused with the similar material being processed (Kroll, 1972; Parkinson et al., 1971, Salzberg et al., 1971).

REFERENCES

Gumenik, W. E., & Rossman, F. A perceptual isolation effect in short-term memory. *Psychonomic Science*, 1970, **19**, 98–99.

Hopkins, R. H., Edwards, R. E., & Gavelek, J. R. Presentation modality as an encoding variable in short-term memory. *Journal of Experimental Psychology*, 1971, **90**, 319–325.

Kroll, N. E. A. Short-term memory and the nature of interference from concurrent shadowing. *Quarterly Journal of Experimental Psychology*, 1972, **24**, 414–419.

Kroll, N. E. A., & Kellicutt, M. R. Short-term recall as a function of covert rehearsal and intervening task. *Journal of Verbal Learning and Verbal Behavior*, 1972, **11**, 196–204.

Parkinson, S. R., Parks, T. E., & Kroll, N. E. A. Visual and auditory short-term memory: Effects of phonemically similar auditory shadow material during the retention interval. *Journal of Experimental Psychology*, 1971, **87**, 274–280.

Parks, T. E., Kroll, N. E. A., Salzberg, P. M., & Parkinson, S. R. Persistence of visual memory as indicated by decision time in a matching task. *Journal of Experimental Psychology*, 1972, **92**, 437–438.

Salzberg, P. M., Parks, T. E., Kroll, N. E. A., & Parkinson, S. R. Retroactive effects of phonemic similarity on short-term recall of visual and auditory stimuli. *Journal of Experimental Psychology*, 1971, **91**, 43–46.

Tversky, B. Pictorial and verbal encoding in a short-term memory task. *Perception & Psychophysics*, 1969, **6**, 225–233.

Wickens, D. D. Encoding categories of words: An empirical approach to meaning. *Psychological Review*, 1970, **77**, 1–15.

Witmer, L. R. The association value of three-place consonant syllables. *Journal of Genetic Psychology*, 1935, **47**, 337–359.

(Received June 15, 1972)

Journal of Experimental Psychology
1973, Vol. 98, No. 1, 138–144

REDUNDANCY AND DIMENSIONAL LOAD IN STIMULUS IDENTIFICATION[1]

ROBERT M. LEVY[2]

Indiana State University

The identification of nine stimuli varying in area and/or brightness was examined in six conditions: two unidimensional, two totally redundant, and two nonredundant. The 36 *S*s, 6 per condition, were run in each of five sessions. The identifiability of the stimuli was independent of redundancy (totally redundant or nonredundant), and it was suggested that the effects of the reduction in dimensional load from the unidimensional and redundant to the nonredundant stimuli may depend on the intradimensional separation of stimulus values. The results were generally consistent with Lockhead's hypothesis that the identifiability of a set of stimuli is a function of their distribution in psychological space.

Lockhead (1970) suggested that the identifiability of a stimulus is a function of the distribution in psychological space of the set of stimuli from which it is selected. The greater the distances among these stimuli in psychological space the greater the identifiability of the stimuli within the set. In testing his model, Lockhead concerned himself primarily with totally redundant multidimensional stimuli.

With values A_1, A_2, A_3, A_4 and B_1, B_2, B_3, B_4 on the two dimensions A and B, we can form a number of redundant sets by pairing one value on A with one and only one value on B. Typically, experimenters have used totally redundant sets that were linearly correlated, e.g., A_1B_1, A_2B_2, A_3B_3, A_4B_4. A different combination, which still satisfies the technical requirements of redundancy, is: A_1B_3, A_2B_1, A_3B_4, A_4B_2. Lockhead (1970) has referred to the second as a "sawtooth" combination. The two sets of redundant stimuli are represented respectively by Xs and Δs in Figure 1; the distribution over the bidimensional discrimination space is greater for the sawtooth set than for the linearly correlated set. As he predicted, Lockhead found increased transmitted information with the sawtooth combination.

Levy and Haggbloom (1971), extending Lockhead's (1970) argument, predicted that the most identifiable set of bidimensional stimuli within the psychological space would be that designated by Os in Figure 1. Since these stimuli are nonredundant, it is impossible for the observer to predict the value on the A dimension, knowing the value on the B dimension. Levy and Haggbloom found that the nonredundant set was the most identifiable, supporting Lockhead's model.

Levy and Haggbloom (1971) pointed out, however, that between the redundant and nonredundant conditions in their experiment the intradimensional spacing or discriminability of stimulus values was confounded with the dimensional load, i.e., the number of values on each dimension. Either factor could lead to increased identifiability. Specifically, whereas there was one step between successive values on each dimension in the redundant conditions, there were three steps ($A_4 - A_1$) between component values used on each dimension in the nonredundant condition. Levy and Haggbloom also suggested that the difference in the dimensional loads between the redundant and nonredundant conditions could have been responsible for their finding. Assume that the size of the stimulus set is fixed at n. In any totally redundant condition (R) the number of values on each dimension A and B is $n_A = n_B = n_R = n$.

---

[1] This research was supported in part by National Science Foundation Grant GB-27668. The efforts of Nikki Tannenbaum and Barbara Weinbaum during data collection are greatly appreciated, as is Gregory Lockhead's critical reading of the manuscript.

[2] Requests for reprints should be sent to Robert M. Levy, Department of Psychology, Indiana State University, Terre Haute, Indiana, 47809.

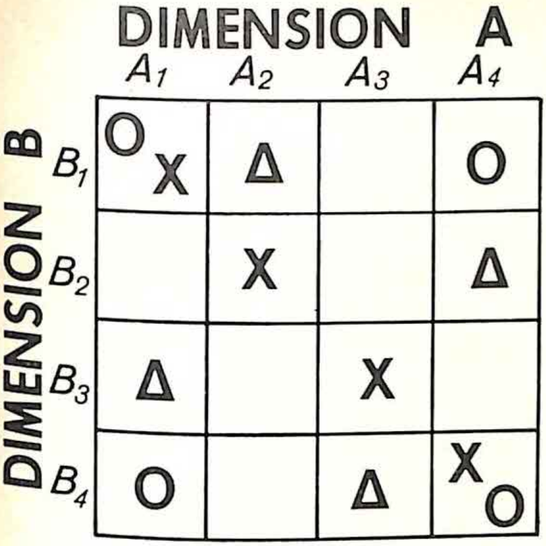


FIG. 1. Distribution of linearly correlated redundant (Xs), sawtooth redundant (Δs), and nonredundant (Os) stimulus sets in hypothetical psychological space.

For example, if n_R is nine, then n_A and n_B are also nine. A nonredundant (NR) set is formed from the orthogonal combination of values on two (or more) dimensions. The size of the nonredundant set, n_{NR}, determines the number of values that are needed on each dimension, n_A and n_B. Then n_A and n_B can be any integers such that $n_A \times n_B = n_{NR}$. (The case of n_A or n_B equal to one is equivalent to unidimensional variation.) For example, if n_{NR} is 9, then n_A and n_B can only be 3; if n_{NR} is 12, there are a number of sets of factors, 2 × 6, 3 × 4, 4 × 3, and 6 × 2.

If man's limited capacity to identify unidimensional stimuli (Miller, 1956) bears on the identification of multidimensional stimuli, then the dimensional load should become an important consideration. With only four stimuli in a condition, it is difficult to evaluate the contribution of the dimensional load since that number is still within our limited capacity. The dimensional load question is more apparent with nine bidimensional stimuli. In redundant conditions there would be nine values on each component dimension, which exceeds the number of stimuli along a single dimension that we can identify; however, in the nonredundant conditions there would be only three values from each dimension. Nye (1962) and Pollack and Ficks (1954), keeping the number of nonredundant stimuli approximately the same, found increased identifiability when the number of dimensions was increased and the dimensional load reduced.

The purpose of the present experiment was to examine the factors of redundancy, defined to include both total redundancy and nonredundancy, and dimensional load in stimulus identification.

METHOD

Subjects

Thirty-six students in introductory psychology at Indiana State University participated in the experiment for extra course credit. Each *S* participated in five $\frac{1}{2}$–$\frac{3}{4}$ hr. experimental sessions on different days. Six *S*s were assigned randomly to each of six conditions.

Experimental Conditions and Apparatus

Stimuli were square patches of Munsell paper varying in area and/or brightness. The composition of the nine stimuli within each of the six experimental conditions is given in Table 1. The nine values of area (A_1, A_2, . . . A_9) were 1.00, 1.15, 1.32, 1.52, 1.75, 2.01, 2.31, 2.66, and 3.06 cm.^{2} and the nine values of brightness (B_1, B_2, . . . B_9) were N4/, 4.5/, 5/, 5.5/, 6/, 6.5/, 7/, 7.5/, and 8/ on the Munsell scale.

Unidimensional conditions. Two unidimensional conditions were included: area, A, and brightness, B. The value of the nonvarying dimension was the middle one of its set.

Redundant conditions. Following Lockhead (1970), redundant linearly correlated (RC) and redundant sawtooth (RS) conditions were selected. The RC condition represented the minimal distribution in the hypothetical psychological space for a totally

TABLE 1

STIMULUS COMPOSITION AND ASSIGNMENT OF RESPONSES TO STIMULI WITHIN EACH CONDITION

Responses	Stimulus conditions					
	B	A	RC	RS	NR 1	NR 4
BUN	B_1A_5	B_5A_1	B_1A_1	B_1A_1	B_4A_4	B_1A_1
BOO	B_2A_5	B_5A_2	B_2A_2	B_4A_2	B_4A_5	B_1A_5
BEE	B_3A_5	B_5A_3	B_3A_3	B_7A_3	B_4A_6	B_1A_9
BOR	B_4A_5	B_5A_4	B_4A_4	B_2A_4	B_5A_4	B_5A_1
BIVE	B_5A_5	B_5A_5	B_5A_5	B_5A_5	B_5A_5	B_5A_5
BIX	B_6A_5	B_5A_6	B_6A_6	B_8A_6	B_5A_6	B_5A_9
BEV	B_7A_5	B_5A_7	B_7A_7	B_3A_7	B_6A_4	B_9A_1
BATE	B_8A_5	B_5A_8	B_8A_8	B_6A_8	B_6A_5	B_9A_5
BINE	B_9A_5	B_5A_9	B_9A_9	B_9A_9	B_6A_6	B_9A_9

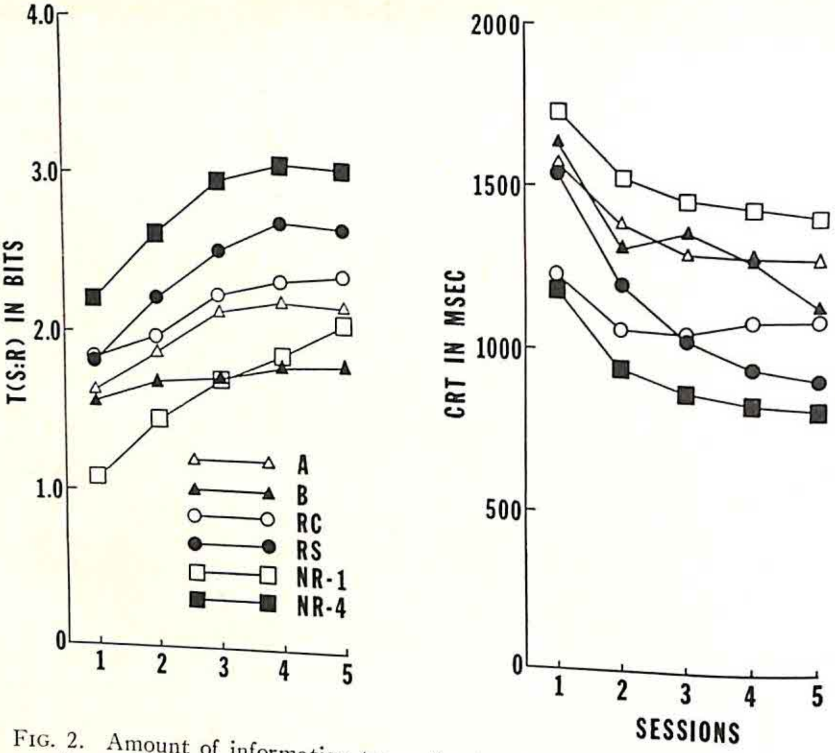


FIG. 2. Amount of information transmitted (left panel) and median choice reaction time (right panel) as functions of experimental condition and session.

redundant condition, whereas the RS condition represented the maximal distribution possible while maintaining the redundancy.

Nonredundant conditions. Within the set of stimulus values selected in the unidimensional conditions, the two nonredundant sets, NR 1 and NR 4, represented the two extremes in terms of distribution in psychological space. The NR 1 set maintained the single-step intraset spacing common to the unidimensional and redundant conditions. The NR 4 set had four steps between the stimulus values on each dimension. For both nonredundant sets, the number of values on each dimension or dimensional load was three compared to the nine values in the unidimensional and redundant conditions.

Stimuli were center mounted on white tagboard 4 × 5 in. (10.16 × 12.70 cm.) for presentation in a Scientific Prototype (Model 800-E) two-channel tachistoscope and subtended on the retina an average angle of 1° 12′. Illumination of the stimulus and pre- and postexposure fields was equated subjectively and set at a level permitting easy viewing of the stimulus. A Hunter millisecond Klockounter was used to measure choice reaction time, and a Hunter noise relay controlled the offset of the stimulus field and clock.

Responses were the numbers ONE, TWO, . . . NINE, with the consonant B prefixed, i.e., BUN, BOO, BEE, BORE, BIVE, BIX, BEV, BATE, BINE. The assignment of responses to stimuli for each condition is also given in Table 1.

Procedure

The procedure was essentially the same one Levy and Haggbloom (1971) used. At the beginning of the first session, *S* was shown the apparatus and the task was described briefly. The nine stimuli for his condition were shown outside the tachistoscope and named. Then the stimuli were shown a few times inside the tachistoscope, and *S* was asked to identify them. When *S* depressed a hand-held push button, the preexposure field of the tachistoscope was switched to the stimulus field and the clock started. The onset of *S*'s verbal response activated the noise relay which switched the tachistoscope's stimulus field to the postexposure field and stopped the clock. Following the trial, *S* was told the correct response and the choice reaction time (CRT). There were 27 practice and 99 test trials in each session; the order of stimulus presentation was randomized with the constraint that the nine stimuli occurred equally often. The *S*s were instructed to respond as quickly and accurately as possible. At the beginning of each of the four remaining sessions the instructions were reviewed.

RESULTS AND DISCUSSION

The amount of information transmitted in bits, T(S:R) (Attneave, 1959), and the median CRT were calculated for each *S* for the 99 test trials in each session and are shown, pooled over *S*s, in Figure 2 as a function of sessions for each condition. Results are reported below, where appropriate, for the fifth session as well as over all sessions.

For T(S:R) the treatment effects, over all sessions—F (5, 30) = 21.3, $p < .001$, fifth session—F (5, 30) = 17.8, $p < .001$; session effects, F (4, 120) = 84.3, $p < .001$; and Treatments × Sessions interaction, F (20, 120) = 2.5, $p < .005$, were significant. Only the session effect, F (4, 120) = 22.2, $p < .001$, was significant for CRT. The large differences shown among the mean CRTs for the different conditions were not significant. Both a nonparametric analysis and an analysis of variance of the fifth-session data showed no significant treatment effects. This result was due to the large within-group variability among the *S*s' median CRTs. (A possible explanation for the variability is presented in a separate section "Subject variability.") The CRT results suggest that gains in T(S:R) were not simply the result of speed–accuracy trade offs. The more accurately identified stimuli were also the more quickly identified. Since the relationships among performance in the various treatment conditions were essentially the same on the different measures, the results are discussed primarily in terms of transmitted information.

Unidimensional conditions. The *S*s transmitted more information with the area stimuli (over all sessions = 2.01 and fifth session = 2.18 bits) than with the brightness stimuli (over all sessions = 1.73 and fifth session = 1.84 bits), over all sessions—F (1, 30) = 4.7, $p < .05$, and fifth session—F (1, 30) = 5.6, $p < .025$. The larger number of just noticeable difference separations for the area stimuli probably led to their increased identifiability, as measured by transmitted information. Since the CRTs for the A condition were not longer than for the B condition, the increase in T(S:R) for area was not gained in a speed–accuracy trade off.

Redundant conditions. Performance in the RS condition (over all sessions = 2.39 and fifth session = 2.68 bits) was better than in the RC condition (over all sessions = 2.16 and fifth session = 2.38); over all sessions—F (1, 30) = 3.2, fifth session—F (1, 30) = 4.2, $p < .05$. Performance was also better in both the RS and RC conditions than in either unidimensional condition. Compared with the unidimensional A condition, approximately .14 bits (over all sessions) and .20 bits (fifth session) more were transmitted in the RC condition, and .38 bits (over all sessions) and .50 bits (fifth session) more in the RS condition.

The consistent gains in identifiability for the RC and RS conditions were not due simply to speed–accuracy trade offs since there were decreases in CRT associated with the increases in T(S:R). Practice was a crucial determinant of the relative identifiability of the RS stimuli. Compared with the RC condition, the CRTs for the RS condition were originally 300 msec. longer than those for the RC condition, becoming approximately 175 msec. shorter on the fifth session; for the RC vs. RS × Sessions interaction, F (4, 120) = 3.5, $p < .025$. The difficulty in identifying the RS stimuli in early sessions may have been due to the lack of a compatible relationship between the stimulus values on both dimensions and the responses.

Biederman and Checkosky (1970) suggested that redundancy gains, on CRT at least, are a function of the degree of overlap of the CRT distributions obtained for the identification of the component unidimensional stimuli. When values are combined on two dimensions in which the identifiability is different, the redundant stimuli should not be identified much more quickly than the more identifiable unidimensional stimuli if there is some overlap of the two CRT distributions, and no more quickly if there is no overlap. The use of an independent group design precludes making some of the comparisons of performance on the unidimensional and redundant conditions necessary to test Biederman and Checkosky's notions. However, the A stimuli used in the present experiment were clearly more identifiable than the B stimuli. The RC and RS stimuli seemed to have been identified more quickly than either set of unidimensional stimuli and more information was transmitted in both redundant conditions than in the A condition. Therefore, although the stimulus values on two dimensions differed in identifiability, the values still combined to form perceptually more distinctive stimuli.

A factor that may determine the form and type of the dimensionality effects on identifiability is the number of stimuli. With two stimuli, as Biederman and Checkosky (1970) used, only correlated redundancy is possible. Moreover, their unidimensional stimuli, although matched on discriminability, were also highly identifiable. For example, their size stimuli increased in area by a factor of 1.7 in one condition and 2.6 in the other. In a two-alternative bidimensional stimulus condition, in which relatively easily discriminable values of size are combined with less discriminable values of brightness, it would seem reasonable to predict that *S*s would respond to the more discriminable, already highly identifiable, stimulus dimension, in which case there would be no redundancy gain, or at least no dimensionality gain. However, in his attempt to process the stimuli efficiently when the information load approaches or exceeds his limited capacity to identify stimuli that vary along a single dimension, *S* might use information from both dimensions, even if the component stimuli differ greatly in identifiability. In this light, Levy and Haggbloom's (1971) failure to find a significant difference in identifiability between their RC and RS conditions may have been due to their use of only four stimuli in each condition.

Nonredundant conditions. The NR 4 stimuli were identified much more accurately than the NR 1 stimuli, over all sessions—F (1, 30) = 76.1, $p < .001$; fifth session—F (1, 30) = 42.9, $p < .001$, and also more quickly, over all sessions—F (1, 30) = 6.9, $p < .025$; fifth session—F (1, 30) = 6.5, $p < .025$. The NR 4 set was the most identifiable of all, as Lockhead's (1970) model predicts. The NR 1 stimuli were initially the least identifiable, but by the fifth session *S*s identified them at a level intermediate to the two unidimensional conditions. (If the CRT data are dropped for one *S* in the NR 1 who deviated greatly from all other *S*s in the condition, there is approximately a 200-msec. decrease in the mean CRT function for the NR 1 condition.)

The expected level of performance in the NR 1 condition, based on the psychological space model, is not as intuitively obvious as it is for the other conditions. Lockhead (1970) states, "the set of stimuli identified more accurately is that one for which the sum of the absolute distances between each stimulus and its immediate neighbors is a maximum and the variance of these distances is a minimum [p. 3]." Since which stimuli constitute "immediate neighbors" and which do not is not self-evident, all interstimulus distances are considered in a crude approximation of the distribution of the stimuli in the discrimination space, assuming that the space is Euclidean, the dimensions A and B are orthogonal, and the adjacent stimulus values represent equal psychological intervals. Then among the k stimuli the $k(k-1)/2$ distances, d_{ij}, are readily computed, $d_{ij} = \sqrt{(A_i - A_j)^2 + (B_i - B_j)^2}$, and from them the mean distance between stimuli, D, and the standard deviation of the distances, σd. Making these assumptions for the A and B conditions in the present study the values of D are: $D_{\mathrm{B}} = 3.33$, $D_{\mathrm{A}} = 3.33$, $D_{\mathrm{RC}} = 4.71$, $D_{\mathrm{RS}} = 5.04$, $D_{\mathrm{NR\,1}} = 1.64$, $D_{\mathrm{NR\,4}} = 6.54$; and the values of σ are: $\sigma\mathrm{B} = \sigma\mathrm{A} = 1.97$, $\sigma\mathrm{RC} = 2.79$, $\sigma\mathrm{RS} = 2.13$, $\sigma\mathrm{NR\,1} = .57$, $\sigma\mathrm{NR\,4} = 2.29$. The estimates of D predict the results reasonably well, except the model predicts that the NR 1 stimuli should be *much less* identifiable than the unidimensional stimuli. The A stimuli were clearly more identifiable than the B stimuli; therefore, the assumption of the same psychological scales for the dimensions would seem inappropriate. Letting the A scale be four times the range of the B scale, the values of D are: $D_{\mathrm{B}} = 3.33$, $D_{\mathrm{A}} = 13.3$, $D_{\mathrm{RC}} = 13.74$, $D_{\mathrm{RS}} = 14.03$, $D_{\mathrm{NR\,1}} = 4.43$, $D_{\mathrm{NR\,4}} = 17.73$; and the values of σ are: $\sigma\mathrm{B} = 1.97$, $\sigma\mathrm{A} = 7.88$, $\sigma\mathrm{RC} = 8.13$, $\sigma\mathrm{RS} = 7.63$, $\sigma\mathrm{NR\,1} = 2.42$, $\sigma\mathrm{NR\,4} = 9.67$. The revised estimates of

D now predict the results in the observed order. Clearly, the intervals on *D* do not correspond to the intervals on T(S:R). Some additional weighting could be included, for example, the variability of the hypothetical interstimulus distances. The identifiability of a set of stimuli could be a joint function, as yet unspecified, of the homogeneity of the interstimulus distance as well as the distance per se. Lockhead's (1970) formulation is essentially the same, except that he measures only the distance between immediate neighbors.

Session effects. That there was significant improvement in performance over sessions on both dependent measures was expected. There was a significant component of the Treatments × Sessions interaction for T(S:R), namely, Sessions × A, B, RC vs. RS, NR 1, NR 4, F (4, 120) = 8.3, $p < .001$. The amount of information transmitted in the RS, NR 1, and NR 4 conditions approached an asymptote more slowly than it did in the A, B, and RC conditions. The finding probably reflects the difference in the S–R relations or compatibility between the two sets of conditions. For A, B, and RC the learned order of the responses corresponded to the natural ordering of the stimuli whereas for the RS, NR 1, and NR 4 conditions this was not possible. It is noteworthy that *S*s did in fact adjust to the particular form of S–R compatability introduced for the latter conditions. The NR 4 stimuli were most identifiable from the outset. Given that the stimulus–response incompatability was an impediment in the first session, then it may well be that the increased separation of the stimulus values on component dimensions was facilitative.

Subject variability. Given the relatively large differences among the mean CRTs for each condition, it was surprising to find that they were not significant. The reason for this was the large error variability. Inspection of the individual *S*'s median CRTs for each session indicated that they were either "short" or "long" for similar amounts of transmitted information. For example, in the fifth session for the RC condition the range of T(S:R) was 2.167–2.601 bits with CRTs of .786, .837, .942, 1.016, 1.288, and 1.637 sec. Although the variability in T(S:R) and CRT may reflect to some degree speed–accuracy trade offs, some other process also seemed to be operating.

The reaction of different *S*s to the instruction "respond as quickly and accurately as possible" may account for the long CRTs. With the exception of the NR 4 condition it was almost impossible for *S*s to be 100% correct, i.e., there was always an element of uncertainty in their decision. It appears that some were reticent to make a decision when they might be wrong. Therefore, they took more time to make their decision although they gained nothing in terms of increased information transmission. The distribution of CRTs in the NR 4 condition support this interpretation; the range of transmitted information was 2.833–3.169 (100% correct) bits and the range of CRTs was .726–.955 sec. In other words, where the identity of the stimulus was certain, *S*s were consistent in their performance.

CONCLUSION

The roles of redundancy and dimensional load in stimulus identification were examined. Defining redundancy to include conditions in which the stimulus values on component dimensions are nonredundant as well as totally redundant, the results of the present experiment suggest that redundancy per se is not an important variable in stimulus identification. Lockhead (1970) showed this with two totally redundant conditions, redundant linearly correlated and redundant sawtooth; Levy and Haggbloom (1971) extended Lockhead's finding to include a nonredundant condition; and in the present experiment, results with two unidimensional, two totally redundant, and two nonredundant conditions support this conclusion.

When the number of stimuli in different conditions is held constant, there is a reduced dimensional load, i.e., number of stimulus values on a dimension, in nonredundant compared with unidimensional and totally redundant conditions. How the reduced dimensional load affects stimulus identification is not clear. For given ranges of stimulus values on component dimensions, the most identifiable stimulus set is the nonredundant one formed by combining the values maximally distributed over the range on each dimension, e.g., the NR 4 condition in the present experiment. Since in this type of condition, the reduction in dimensional load allows the selection of stimulus values on component dimensions that are further apart, the effects of dimensional load and intradimensional separation are confounded. The importance of the intradimensional separation was shown by the superior identifiability of the NR 4 compared with the NR 1 condition in which the intradimensional separation was the same as in the other conditions.

To determine the extent to which the particular combination of dimensions, for example, whether or not it is integral (Garner & Felfoldy, 1970; Lockhead, 1966), affects how stimuli are identified will require research with different dimensional combinations. For the dimensions of area and brightness used in the present experiment, Lockhead's (1970) model seems to offer the basis for a parsimonious explanation which indicates a primary role for dimensionality per se and the distribution of stimuli in a psychological space and only a secondary role for the presence or absence of redundancy in multidimensional stimulus identification.

REFERENCES

Attneave, F. *Applications of information theory to psychology.* New York: Holt, Rinehart & Winston, 1959.

Biederman, I., & Checkosky, S. F. Processing redundant information. *Journal of Experimental Psychology*, 1970, **83**, 486–490.

Garner, W. R., & Felfoldy, G. L. Integrality of stimulus dimensions in various types of information processing. *Cognitive Psychology*, 1970, **1**, 225–241.

Levy, R. M., & Haggbloom, S. J. Test of a multidimensional model of stimulus identification. *Psychonomic Science*, 1971, **25**, 203–204.

Lockhead, G. R. Effects of dimensional redundancy on visual discrimination. *Journal of Experimental Psychology*, 1966, **72**, 95–104.

Lockhead, G. R. Identification and the form of multidimensional discrimination space. *Journal of Experimental Psychology*, 1970, **85**, 1–10.

Miller, G. A. The magical number seven plus or minus two. *Psychological Review*, 1956, **63**, 81–97.

Nye, P. W. Aural recognition time for multidimensional signals. *Nature*, 1962, **196**, 1282–1283.

Pollack, I., & Ficks, L. Information of elementary multidimensional auditory displays. *Journal of the Acoustical Society of America*, 1954, **26**, 155–158.

(Received June 15, 1972)

Journal of Experimental Psychology
1973, Vol. 98, No. 1, 145–151

EFFECT OF EXPLICIT TRIAL-BY-TRIAL INFORMATION ABOUT SHOCK PROBABILITY IN LONG INTERSTIMULUS INTERVAL GSR CONDITIONING[1]

ARNE ÖHMAN,[2] PÄR-ÅKE BJÖRKSTRAND, AND PER-ERIK ELLSTRÖM

University of Uppsala, Uppsala, Sweden

Stimuli giving explicit information about probabilities of 0%, 25%, 50%, 75%, and 100% of receiving a strong or a weak shock were presented for 32 *S*s in a conditioning paradigm with an 8-sec. interstimulus interval. Number of shocks were matched to probability over four replications of each stimulus. CS and pre-UCS responses showed monotonically increasing trends in magnitude with probability whereas post-UCS and unconditioned responses gave inverted-**U** and **U**-shaped relationships, respectively. The strong shock condition produced bigger CS, pre-UCS, and unconditioned responses.

Two types of relationships between probability of a future aversive event and anticipatory activation have been suggested. Grings (1960, 1969) represents the first line of thought, implying that higher probability of an aversive stimulus leads to a firmer "preparatory set," which in turn produces larger anticipatory responses. Furthermore, the set modifies the responses to the expected stimulus, giving rise to an inhibitory effect if the expectancy is confirmed, and to a disinhibitory effect, i.e., a perceptual disparity response (PDR), if the expectancy is disconfirmed.

Epstein (1972; Epstein & Roupenian, 1970) represents the second approach, emphasizing uncertainty as central to the understanding of anticipatory activation. The uncertainty produced by a failure to formulate an expectancy regarding future stimulation is specified as a primary determinant of arousal, and therefore, an inverted-**U** relationship between anticipatory activation and probability of an aversive stimulus is predicted.

The empirical data regarding this problem are quite scanty. Niemelä (1969) obtained support for Grings' (1969) conception, observing a monotonically rising gradient with increased probability of electric shock both for the largest response and for frequency of skin conductance responses (SCRs) during 20-sec. anticipation in a modified, continuous count-up procedure.

Epstein and Roupenian (1970) presented three groups of *S*s with probabilities of 5%, 50%, and 95% of receiving shock at the end of a discrete auditory count up. They found higher skin conductance level (SCL) and larger SCRs during anticipation in the 5% than in the 50% and 95% conditions. Interpreting these results in terms of an inverted-**U** relationship, they assumed a discrepancy between objective and subjective probability, leading to maximal subjective uncertainty in the 5% condition. However, lacking definite independent information regarding subjective shock probabilities, this interpretation remains of limited value.

From these studies it may be concluded that the data do not unequivocally support either of the theoretical conceptions, although they seem to be more in line with Grings' (1969) than with Epstein's (1972) notion.

The aim of the present study was to pursue this analysis a bit further by exploring the effect of explicit information about shock probability in a long interstimulus interval (ISI) GSR conditioning paradigm, permitting the observation of multiple responses to the CS–UCS complex (Lockhart, 1966). Such multiple responses have previously been demonstrated to reflect different psychological processes (Öhman, 1971a, 1971b, 1972). Thus, the effect of five CSs signaling different probabilities of two levels of UCS intensity was studied.

---

[1] This research was supported by a grant to the first author from the Social Science Research Council of Sweden. The authors acknowledge their gratitude to Ulla Andersson and Ingrid Wirenstrand for help in collecting the data.

[2] Requests for reprints should be sent to Arne Öhman, Department of Psychology, Slottsgränd 3, S-752 20, Uppsala, Sweden.

METHOD

Subjects. Thirty-two undergraduate psychology students at the University of Uppsala, 15 female and 17 male, were paid to participate in the experiment.

Apparatus. Palmar skin resistance was recorded on a Servogor 1-channel inkwriter (Type RE 511, Goerz) through a dc constant-current bridge circuit applying a current magnitude of 11 μa. Beckman silver/silver chloride electrodes, 8 mm. in diameter, embraced in a plastic cup, were used. The electrodes were attached to the first phalanx of the first and second fingers of the nondominant hand by adhesive electrode collars. Beckman-Offner electrode paste was used as contact medium. Electric shocks (UCS) were delivered through silver electrodes attached to the volar surface of the first and second fingers of the dominant hand. The shocks were produced by a stabilized dc current that charged a capacitor and whose voltage could be manipulated by *E*. Upon closing a contact the capacitor discharged over a transformer, giving a maximal output of 660 v. at a resistance of 10 kΩ. The duration of the shock was .05 sec., inducing a sharp, pin-pricking sensation. Dia-slides (24 × 35 mm.) were used to project the CSs on a milk-glass screen, 2 m. in front of the *S*. The stimuli were produced by a Sawyer projector, the duration being controlled by an electronic timer, and the intertrial interval by a Tandberg tape recorder activating relay detectors. The apparatus was connected to the power line through a voltage stabilizer.

The CSs as projected on the screen were 45 × 80 cm. with text of 3.5 cm. height in the center, indicating the probability for and the intensity of the potential UCS (e.g., "25% STRONG"). The stimulus intensity was about 45 lx. (background) to 65 lx. (text).

Design. A 2 × 5 factorial design with repeated measurement was used. The independent variables were UCS intensity and objective probability of UCS. For the former, strong and weak shocks were used; for the latter, five probabilities of shock: 0%, 25%, 50%, 75%, and 100%.

Experimental procedure. When the *S* entered the laboratory, he was seated in a comfortable chair, the skin of the electrode sites were cleaned with distilled water, and the electrodes were applied as described above. The procedure of the experiment proper was then explained to the *S*. He was told about the probabilities of getting a shock and about the two shock intensities. It was emphasized that the probabilities shown on the screen were "true" in the sense that the figures given reflected a correct proportion of shocks to be delivered within probability during the experiment. To obtain the *S*'s own estimates of probability, he was asked to state a "yes" or "no" hypothesis after each trial, depending on whether or not he had believed a shock would be delivered, independent of whether shock actually was given. The estimation had to be given after trials, not to interfere with the response pattern to the CS.

The shock level was set individually, with the instruction that the shock was to be perceived as "clearly unpleasant, but not painful." The strong shock was set 25% above the level chosen by the *S*; the weak shock, 25% below.

Four replications of each of the 10 CSs were made, one in each block of 10 trials. The CS order within blocks was randomized, so that 16 different presentation orders were obtained, i.e., two *S*s for each order of presentation. The reinforced trials within probability were randomized independently between the blocks for each *S*. The CS duration was 8 sec., with the UCS, if delivered, immediately following CS offset. Forty trials at a mean trial interval of 30 sec. ± 10 sec. were presented.

Analysis of data. CS, pre-UCS, post-UCS, and unconditioned responses were defined in terms of response latency according to the suggestions by Lockhart (1966). The CS response was scored 1–4 sec. after CS onset, the pre-UCS response from 4 sec. after CS onset to 1 sec. after CS offset, and the post-UCS and the UCR 1–4 sec. after CS offset. If more than one response occurred during the pre-UCS interval, the more frequently found component was scored. The responses were transformed into log micro-mho units to obtain data expressed as change in log conductance. The post-UCS and UC responses were blocked within UCS intensity and probability. Thus, the post-UCS response for "50% STRONG" was based upon two responses; the "25% WEAK" upon three responses, etc.

RESULTS

CS responses. Figure 1 shows the CS response as a function of shock probability for the four trial blocks. Analysis of variance revealed significant main effects of probability, F (4, 124) = 6.30, $p < .001$; intensity, F (1, 31) = 17.79, $p < .001$; and trials, F (3, 93) = 4.05, $p < .01$. The Probability × Intensity interaction was significant, F (4, 124) = 4.29, $p < .001$, due to diverging trends over probability for the two intensities. Follow-up analyses with the Newman-Keuls procedure yielded no significant differences for the weak intensity, whereas for the strong intensity, the 100% condition exceeded all the others, which did not differ between themselves. The Probability × Intensity × Trials interaction was also significant, F (12, 372) = 2.44, $p < .01$. As can be seen in Figure

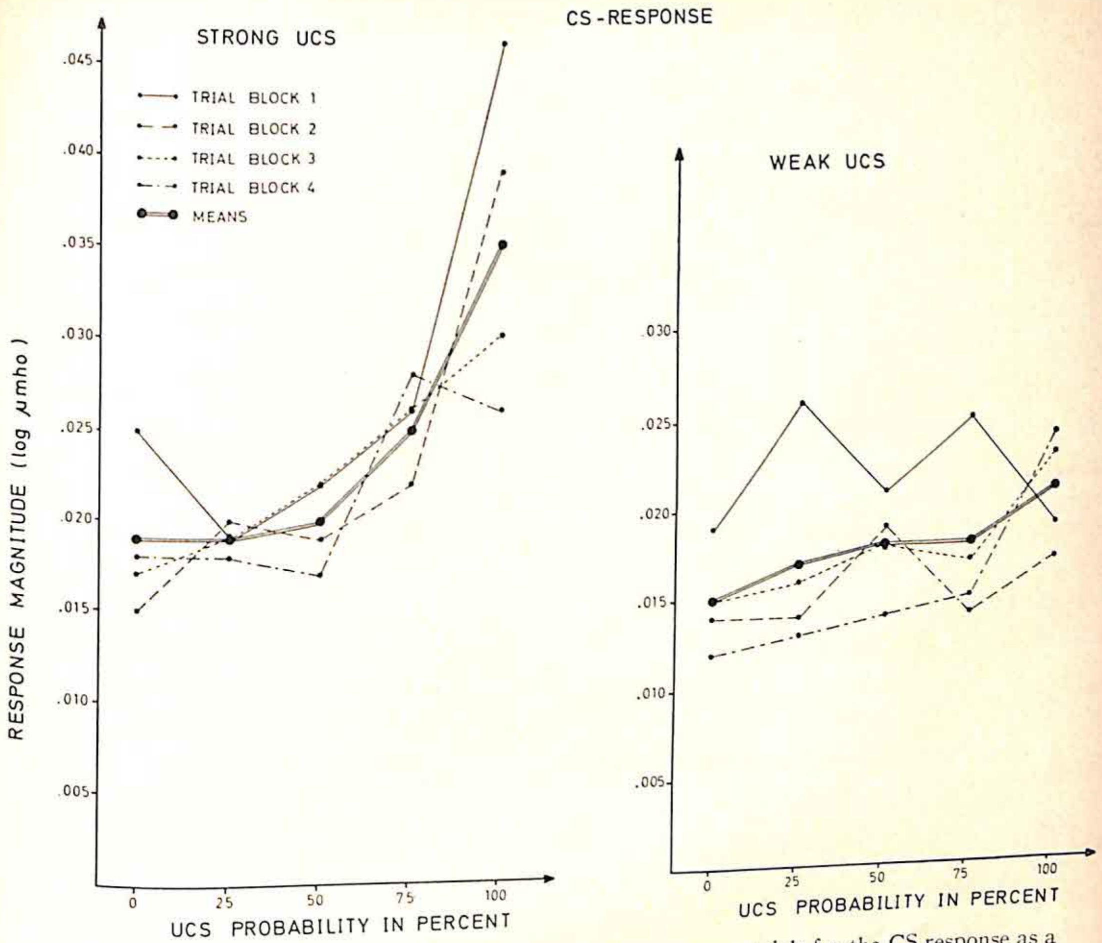


FIG. 1. Mean response magnitude and response development over trials for the CS response as a function of UCS probability. (Left diagram: strong UCS. Right diagram: weak UCS.)

1, this effect pertained to a successive flattening out of the strong intensity probability gradients over trials, whereas the weak intensity gradients tended to increase their slopes over trials.

Pre-UCS responses. Figure 2 shows the pre-UCS responses as a function of probability and trials. All the main effects were significant, F (4, 124) = 7.74, $p < .001$; F (1, 31) = 10.51, $p < .01$; and F (3, 93) = 12.12, $p < .001$, for probability, intensity, and trials, respectively. None of the interactions was significant. Newman-Keuls tests showed the 100% and 75% conditions to differ from the 0% condition, but not between themselves, nor from the 25% and 50% for the strong intensity. For the weak intensity, the 100% condition differed from the 0% condition.

Post-UCS responses. As shown in Figure 3, the post-UCS responses demonstrated an inverted-**U** relation with probability. The only significant effect pertained to probability, F (3, 93) = 18.62, $p < .001$. Follow-up tests showed that the 0% condition was significantly smaller than the other three conditions, and that both 25% and 50% exceeded the 75% condition.

UCR. The UCRs showed a **U**-shaped relationship to probability, as can be seen in Figure 3. The main effects of probability, F (3, 93) = 4.07, $p < 01$, and intensity, F (1, 31) = 6.91, $p < .05$, were significant. Follow-up tests showed the 25% condition to exceed the 50% and 75% conditions, but not the 100% condition. The latter treatment, however, did not differ from the two former treatments.

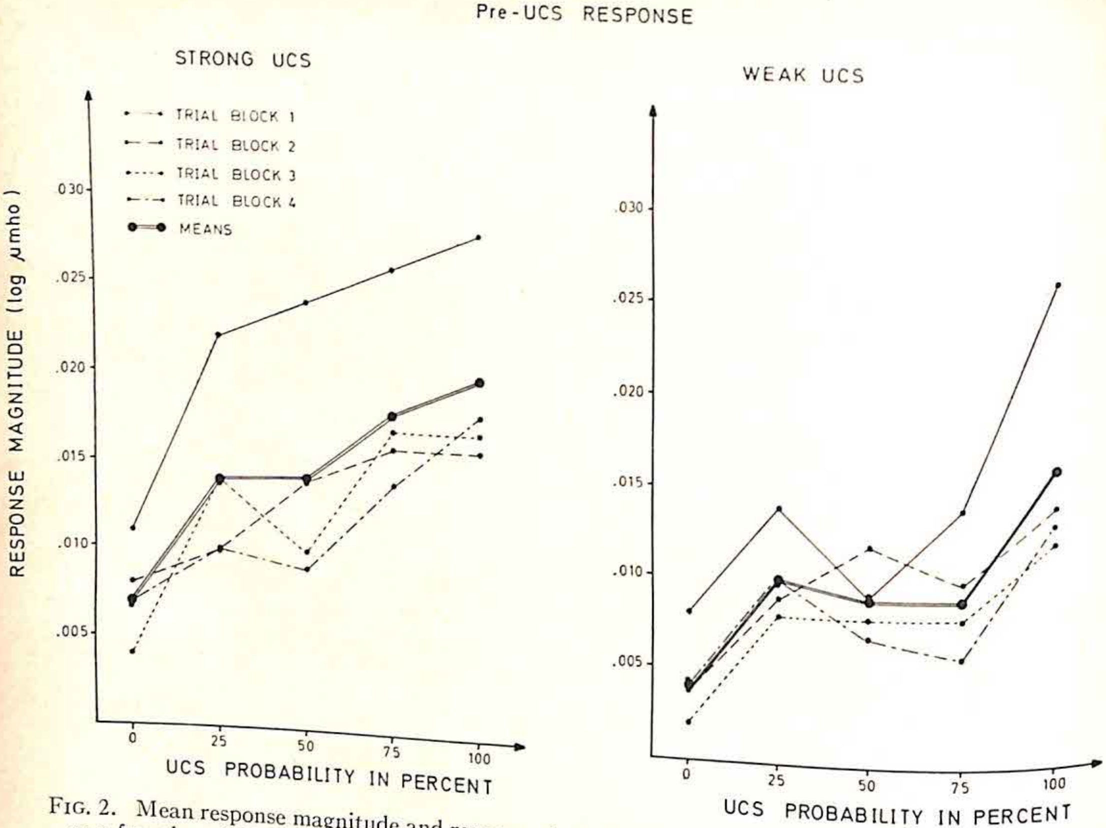


FIG. 2. Mean response magnitude and response development over trials for the pre-UCS response as a function of UCS probability. (Left diagram: strong UCS. Right diagram: weak UCS.)

Subjective probabilities. Table 1 shows the subjective probabilities for the two intensity levels as a function of objective probability. The effect of probability, of course, was significant, F (4, 124) = 241.02, $p < .001$. More interesting, there was a significant tendency to judge shock as more probable with high than with low intensity, F (1, 31) = 13.66, $p < .001$.

Response magnitude as a function of subjective expectancy. To assess the effect of subjective expectancy on response magnitude, the first instance of a "yes" response in the 50% condition was compared to the first instance of a "no" response, irrespective of shock intensity. Table 2 shows the results. Similar results were obtained when the analysis was based on the mean of all observations. As can be seen, there were tendencies toward differences for the pre- and post-UCS responses. However, none of the differences reached statistical significance as assessed by t tests.

DISCUSSION

After the completion of the present experiment, Grings and Sukoneck (1971) reported data from a quite similar experimental situation. The results will, therefore, be discussed in relation to their findings.

Both studies demonstrated monotonically increasing CS responses with increased probability of shock. This effect, however, seems to be somewhat elusive. In the present experiment, the significant effect of probability was mainly attributable to the strong shock condition, whereas no significant effect was observed for the weak shock condition. Furthermore, as indicated by the three-way interaction, the effect of probability in the strong shock condition was most clearly manifested in the first trial block, tending to disappear with training due to more pronounced habituation in the 100% condition. Similarly, Grings and Sukoneck (1971), who used only one UCS intensity, found that the significant effect observed early in training disappeared after 40 trials.

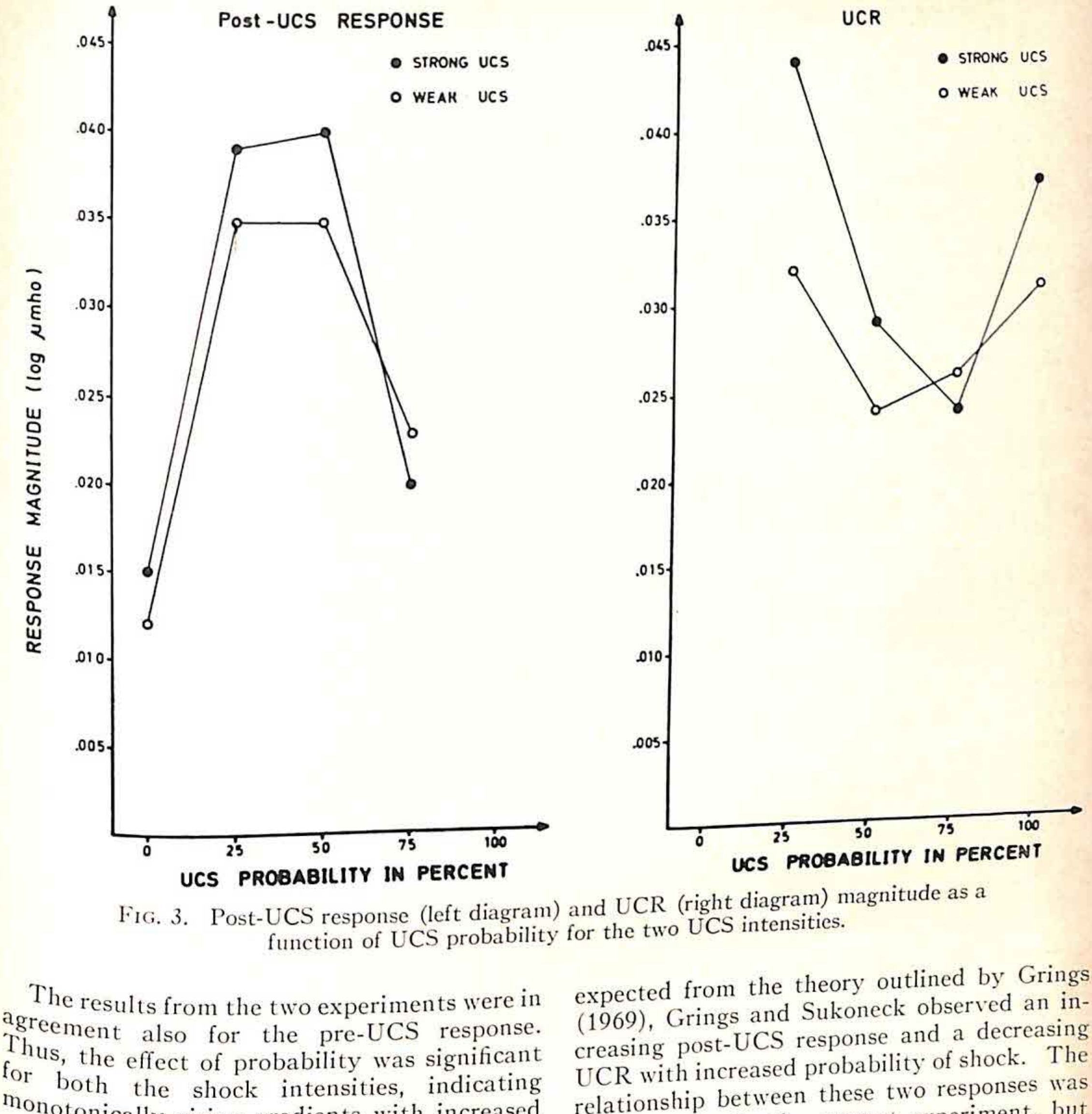


FIG. 3. Post-UCS response (left diagram) and UCR (right diagram) magnitude as a function of UCS probability for the two UCS intensities.

The results from the two experiments were in agreement also for the pre-UCS response. Thus, the effect of probability was significant for both the shock intensities, indicating monotonically rising gradients with increased probabilities, as was reported by Grings and Sukoneck (1971).

For the post-UCS response and the UCR, however, the results from the present experiment are different from those of Grings and Sukoneck. In accordance with what could be expected from the theory outlined by Grings (1969), Grings and Sukoneck observed an increasing post-UCS response and a decreasing UCR with increased probability of shock. The relationship between these two responses was inverted also in the present experiment, but the post-UCS response showed an inverted **U**-shaped function, and the UCR a **U**-shaped relationship to probability of shock.

A number of procedural differences might contribute to the discrepant results for the

TABLE 1

MEAN PROPORTION OF "YES" HYPOTHESES (IN PERCENTAGES) AS A FUNCTION OF OBJECTIVE PROBABILITY AND UCS INTENSITY

UCS intensity	Objective probability				
	0	25	50	75	100
Weak	0.0	20.3	56.3	91.4	96.1
Strong	3.9	28.1	71.1	96.1	98.4

TABLE 2

RESPONSE MAGNITUDE AS A FUNCTION OF SUBJECTIVE EXPECTANCY

Expectancy	Type of response			
	CS	Pre-UCS	Post-UCS	UCR
"Yes"	.023	.014	.029	.046
"No"	.023	.009	.023	.044
n	28	28	20	21

post-UCS response and the UCR in the two studies. First, Grings and Sukoneck did not include a 50% condition in their study. Since this condition involves maximal uncertainty, it might be critical, and the conclusions regarding the form of the relationship are complicated lacking the 50% point. Second, and probably more important, their *S*s had to learn the contingencies between different CSs and shock probabilities through experience. It is an interesting possibility that expectancies developed from actual experience could differ in their effects from expectancies instigated through instructions (Epstein, 1972; Skinner, 1969, p. 115). A third difference in procedure of potential importance pertains to the method of stating subjective hypotheses about shock. Thus, the *S*s of Grings and Sukoneck made their hypothesis explicit before the reception of the UCS, which introduces a component of self-appraisal in the situation (cf. Epstein & Bahm, 1971), whereas the *S*s in the present experiment reported their expectancy after reception of the UCS.

The discrepant results of these two experiments indicate that the effects for the post-UCS responses and the UCRs might be dependent on quite narrowly defined experimental conditions. Further support for such a conclusion was obtained in an unpublished experiment where 32 *S*s received one reinforced and one unreinforced presentation of CSs indicating 5%, 25%, 50%, 75%, and 95% probability of shock, while being led to believe that probabilities should match number of shocks during a half-hour experimental session. There were tendencies toward a probability effect for the CS and pre-UCS responses, but significantly so only for the latter. The UCR showed a nonsignificant U-shaped relationship, while the post-UCS response showed an essentially flat function, with only a slight elevation at 50%. Thus, whereas the results were similar to those of the present experiment, they seem to imply that consonance between probability and number of shocks is a factor behind the significant results. The significant effect of probability on the pre-UCS response is in agreement with previous findings of this measure as especially sensitive to cognitive influences (Fuhrer & Baer, 1969; Lockhart, 1966; Öhman, 1971b).

In conclusion, then, the CS and especially the pre-UCS response seem to show the relationship to shock probability required by Grings' (1969) preparatory set theory. For the post-UCS response and the UCR, however, this seems to hold only under certain conditions. Therefore, anticipatory, perceptual disparity and shock responses cannot be taken as equivalent indicators of a unitary preparatory set (cf. Öhman, 1972).

Comparing the subjective probabilities stated by the *S*s in the present experiment with the objective ones, it is clear (*a*) that there was a tendency to overestimate the shock probabilities, especially the intermediate ones, and (*b*) that this tendency was more pronounced for the strong shock condition. Thus, a more aversive event seemed to induce a more pessimistic hypothesis regarding the probability of experiencing it than a less aversive event.

The latter finding might imply that the effect of shock intensity on response magnitude was not an effect of intensity per se, but was mediated by the higher probabilities ascribed to the high-intensity shock. However, if CS and pre-UCS response magnitudes for the two intensities are plotted against subjective probabilities, the intensity effect becomes somewhat attenuated, but the points still seem to retain characteristics of two different intensity functions.

The present study failed to confirm the direct relationship between subjective expectancy and magnitude of response reported by Epstein and Bahm (1971). The procedure used by these authors, however, might involve a confounding between verbal responses and subjective expectancy as elicitors of GSRs, making their results difficult to interpret. The procedure of the present study, on the other hand, might have been quite insensitive since it employed a "forced-choice" retrospective assessment. Thus, the reported tendencies (see Table 2) might suggest the possibility of obtaining effects if a more sensitive procedure of measurement could be devised.

REFERENCES

EPSTEIN, S. The nature of anxiety with emphasis upon its relationship to expectancy. In C. D. Spielberger (Ed.), *Anxiety: Current trends in theory and research.* Vol. 2. New York: Academic Press, 1972.

EPSTEIN, S., & BAHM, R. Verbal hypothesis formulation during classical conditioning of the GSR. *Journal of Experimental Psychology*, 1971, **87**, 187–197.

EPSTEIN, S., & ROUPENIAN, A. Heart rate and skin conductance during experimentally induced anxiety: The effect of uncertainty about receiving a noxious stimulus. *Journal of Personality and Social Psychology*, 1970, **16**, 20–28.

Fuhrer, M. J., & Baer, P. E. Cognitive processes in differential GSR conditioning: Effects of a masking task. *American Journal of Psychology*, 1969, **82**, 168–180.

Grings, W. W. Preparatory set variables related to classical conditioning of autonomic responses. *Psychological Review*, 1960, **67**, 243–250.

Grings, W. W. Anticipatory and preparatory electrodermal behavior in paired stimulus situations. *Psychophysiology*, 1969, **5**, 597–611.

Grings, W. W., & Sukoneck, H. I. Prediction probability as a determiner of anticipatory and preparatory electrodermal behavior. *Journal of Experimental Psychology*, 1971, **91**, 310–317.

Lockhart, R. A. Comments regarding multiple response phenomena in long interstimulus interval conditioning. *Psychophysiology*, 1966, **3**, 108–114.

Niemelä, P. Electrodermal responses as a function of quantified threat. *Scandinavian Journal of Psychology*, 1969, **10**, 49–56.

Öhman, A. Differentiation of conditioned and orienting response components in electrodermal conditioning. *Psychophysiology*, 1971, **8**, 7–22. (a)

Öhman, A. Interaction between instruction-induced expectancy and strength of unconditioned stimulus in GSR conditioning. *Journal of Experimental Psychology*, 1971, **88**, 384–390. (b)

Öhman, A. Factor analytically derived components of orienting, defensive and conditioned behavior in electrodermal conditioning. *Psychophysiology*, 1972, **9**, 199–209.

Skinner, B. F. *Contingencies of reinforcement.* New York: Appleton-Century-Crofts, 1969.

(Received June 15, 1972)

Journal of Experimental Psychology
1973, Vol. 98, No. 1, 152–159

TRANSFER OF RESPONSE GROUPINGS IN PAIRED-ASSOCIATE LEARNING[1]

PEGGY A. RUNQUIST[2]

University of Alberta, Edmonton, Alberta, Canada

The role of response-term organization in paired-associate learning was investigated by requiring *S*s to learn two paired-associate lists in which the stimulus terms were grouped into conceptual categories. In the second list, responses were either re-paired within the same category, shifted to a different category, or re-paired with instances of different categories. Other groups had the same categories in List 2, but different instances as stimuli. List 2 learning was followed by free recall of the responses. Increased within-category errors and significant clustering of responses according to stimulus categories were found with within-category re-pairing, but not when responses were re-paired with another category or across categories. The result held even when stimuli were changed. These results indicate that responses are not organized per se, but that response grouping is mediated by stimulus categories.

This experiment investigated the process of response-term grouping in paired-associate (PA) learning by which the response terms are subcategorized into sets corresponding to similarity groupings among the stimulus terms. The two questions to which the research was addressed were: (*a*) What is the nature of the relationship among the responses that are grouped together? (*b*) What is the effect of this grouping on subsequent acquisition performance?

Response-term grouping has been described as a stage occurring prior to associative learning, during which *S* learns to categorize the response terms into sets of possible correct responses to corresponding sets of similar stimuli. The *S*s at this stage have presumably learned that the response terms, R_1, R_2, and R_3 are appropriate alternative responses for the set of similar stimuli, S, S′, and S″, but have not yet learned specific associations within this set.

Support for this grouping stage has come primarily from two findings: (*a*) Overt errors in learning paired-associate lists with groupable similar stimuli tend to be predominately confusion errors from within a similarity set (e.g., Joinson & Runquist, 1968; Runquist, 1970; Runquist, 1966, 1968; Underwood, Ekstrand, & Keppel, 1965, Experiment V), and (*b*) response terms paired with sets of similar stimuli tend to be recalled as clusters on a free-recall test following paired-associate learning (Runquist, 1970). Such findings, however, do not necessarily implicate grouping to the exclusion of other processes. The predominance of confusion errors from within sets of similar items, for example, could be explained simply on the basis of generalization gradients, i.e., associative strength being built up via stimulus generalization between a response and a stimulus similar to the one with which it is paired, but not between a response and a dissimilar stimulus. If this were the case, these errors would be more a reflection of stimulus confusion than of response categorization.

The free recall of responses in clusters based on similarity groupings among the stimuli can also be accounted for without postulating a separate response-grouping stage. If, for example, free recall of the responses occurred by first recalling specific stimulus items and then retrieving response information associated with each stimulus, clustering could simply reflect (*a*) the existence of generalized error tendencies remaining at the end of acquisition, or even (*b*) the organization of the stimulus terms themselves at the time of recall. In the absence of conclusive evidence about the mechanism(s) responsible for these two findings, even the existence of response grouping as a stage in learning is questionable.

---

[1] This experiment is based on a dissertation submitted to the University of Alberta in partial fulfillment of the requirements for the PhD degree. The research was conducted under the supervision of W. W. Rozeboom and was supported in part by Research Grant APA-88 from the National Research Council of Canada to W. N. Runquist.

[2] Requests for reprints should be sent to Peggy A. Runquist, Department of Psychology, University of Alberta, Edmonton, Alberta, Canada.

The present study was designed in part to provide evidence bearing on this issue. If the *S* does learn to categorize the responses into sets, these response groupings should transfer to a second list in much the same manner as do specific S–R associations, i.e., grouping should be increased if the relationship among the grouped responses is the same in the second list as in the first list and decreased if the relationship is different. In this experiment, *S*s learned two categorized PA lists (categorized stimuli, unrelated responses), conforming to the A–B, A–B_R transfer paradigm, and then recalled the response terms on a free-recall test following second-list learning. Three types of S–R re-pairing were used. In within-category (WC) re-pairing, the response terms paired with stimuli from a given category in the first list were re-paired with stimuli from the same category in the second list. In between-category (BC) re-pairing, these responses were re-paired with the stimuli from a different first-list category. Finally, in across-category (AC) re-pairing, these responses were re-paired each with a stimulus item from a different first-list category. Each of these re-pairing schemes was used to disrupt the various types of relationships that could exist among the responses grouped together during first-list learning which could account for the occurrence of within-category confusion errors and response-term clustering. Three types of relationships were considered: (*a*) common associations with specific stimulus items, perhaps as a result of stimulus generalization; (*b*) common associations with a specific conceptual category represented by the similar stimuli; and (*c*) direct associations among the responses themselves formed as a result of common associations either with specific stimulus items or with stimulus categories. With WC re-pairing, all three types of relationships are maintained in the second list; hence second-list grouping should be increased. With BC re-pairing, both the first and second types of relationship among the responses would be disrupted leaving only direct interresponse associations intact. If either specific stimulus items or specific stimulus categories are necessary for maintaining response groupings, BC re-pairing should result in decreased grouping. On the other hand, if direct interresponse associations exist, increased grouping should occur. With AC re-pairing, decreased grouping is to be expected since all three relationships are different in the second list.

Although the three re-pairing arrangements with identical stimuli do allow one to distinguish between a direct interresponse association mechanism and a common association mechanism for explaining response clustering and within-category confusion errors, the three arrangements do not separate common associations with specific stimulus items from common associations with a stimulus category. The three re-pairing schemes were repeated using different instances of the same categories as stimuli in the second list in order to determine if specific stimulus items are necessary for groupings to be maintained or whether stimulus categories are sufficient. An additional two conditions with completely new stimulus categories in the second list and either BC or AC response re-pairing were also included. Evidence of increased grouping in the first of these conditions would provide additional support for direct interresponse associations. In the latter condition, increased grouping would represent nonspecific grouping transfer or a set to group responses on the basis of stimulus characteristics.

The second purpose of the experiment was to assess the effect of grouping on acquisition performance. It has been suggested (Battig, 1966, 1968; Runquist, 1966, 1968) that not only does grouping occur, but that it is a primary means by which interpair interference can be reduced. According to this view, in learning a single pair grouping restricts the interference to other items from within the same subset. Facilitation could occur as a result of a reduction in the amount of overall interference or as a result of guessing, if *S* were to learn the responses to some of the stimuli within a set and respond to the others correctly on the basis of elimination (Joinson & Runquist, 1968; Runquist, 1966). Regardless of the mechanism by which facilitation might occur, there is surprisingly little evidence that it does occur—at least when stimulus similarity forms the basis for the grouping. Significantly better performance with groupable high-similarity lists has been found once in a transfer task (Runquist, 1968) and once in a retention task (Joinson & Runquist, 1968). Support has been largely indirect in other studies, coming from the finding of no significant difference between a high-similarity categorizable list and a noncategorizable low-similarity list, the former, by virtue of its similarity, being expected to provide considerable interference (e.g., Runquist, 1970; Runquist, 1966).

In the present experiment, it was expected that if the categorization of responses into sets did facilitate performance, second-list learning would proceed more rapidly if response groupings in the transfer task were the same as in the first task than if they were new. More specifically, conditions which lead to increased response grouping, as indicated by an increase in within-category confusion errors and response clustering, would also be expected to produce faster learning relative to other re-pairing conditions having the same interlist stimulus similarity.

METHOD

Design

There were eight basic transfer conditions in the experiment. Labeled on the basis of the relation of the first to the second (A–B) list, the conditions were: A–B_{WC}, A–B_{BC}, A–B_{AC}, A′–B_{WC}, A′–B_{BC}, A′–B_{AC}, C–B_{BC}, and C–B_{AC}. The first letter indicates the similarity of the stimulus terms in the two lists to one another: Stimuli were identical (A), different instances of the same conceptual categories (A′), or instances of different conceptual categories (C). The remaining letters indicate three different S–R re-pairing schemes of the same (B) responses. The response paired with a given category of three stimuli in the second list was re-paired for the first list with stimuli from the same category (WC or within-category re-pairing), with the three stimuli from a different category (BC or between-category re-pairing), or with three stimuli from three different categories (AC or across-category re-pairing). In addition, there was one control group (A–B) that learned only the second list.

With two sets of lists used under each of the conditions, there was a total of 18 groups.

Materials

Two identically constructed sets of 16 12-item paired-associate lists were used as materials. Each list consisted of three nouns from four different conceptual categories as stimulus terms and 12 unrelated adjectives as response terms. The nouns were selected from the 12 most frequent responses to category names as determined by the Battig and Montague (1969) norms, and were assigned to lists in such a way that mean taxonomic frequency was approximately equal for all lists. The adjectives were two- or three-syllable words of A or AA frequency of occurrence according to the Thorndike-Lorge G count.

All stimulus terms for a given set of lists were selected from a pool of six nouns from each of eight different conceptual categories. The lists were so constructed that the same four categories appeared in half (eight) of the lists, with each category being represented in four lists by one group of three nouns, and in the other four lists by the remaining three nouns. These four basic lists of nouns were used as stimulus terms for first and second lists equally often under all conditions. Each list of stimuli was paired in four different ways with the same 12 adjectives. Because it is not possible to completely counterbalance S–R pairings with categorized lists (Zavortink & Keppel, 1968), a basic pairing for the second (A–B) list was selected for each list of nouns, taking care that in lists containing different instances of the same categories as stimuli, the same three adjectives were associated with these nouns. The remaining S–R pairings used in the first lists were then generated from the A–B lists by re-pairing responses within, between, or across categories, as described earlier. In constructing the specific pairings, care was taken to avoid obvious associations between the stimulus and response members and among the responses paired with stimuli from a given category.

Subjects

The *S*s were 216 students from an introductory psychology course, serving in order to obtain course credit. The *S*s were assigned to the nine conditions in order of their appearance at the laboratory according to a prearranged scheme randomizing the order of conditions within blocks containing the four sets of stimuli used under each condition. The two sets of lists were run consecutively with 12 *S*s

TABLE 1

Acquisition Measures for List 1 and List 2 Learning

Measure	Condition								
	A-B	A-B$_{WC}$	A-B$_{BC}$	A-B$_{AC}$	A′-B$_{WC}$	A′-B$_{BC}$	A′-B$_{AC}$	C-B$_{BC}$	C-B$_{AC}$
List 1									
Trials to criterion	—	9.96	11.17	9.46	10.83	9.88	9.71	9.42	8.46
Proportion of within-category errors	—	.43	.58	.44	.59	.50	.53	.60	.63
List 2									
Errors on Trial 1	9.00	9.00	9.25	9.33	6.96	8.08	8.25	6.33	6.79
Proportion of within-category errors on Trial 1	.27	.83	.25	.18	.83	.12	.13	.34	.40
Trials to criterion	11.70	12.54	12.17	11.00	7.88	8.46	8.33	6.67	5.96
Proportion of within-category errors to criterion	.47	.76	.32	.30	.73	.34	.41	.52	.39

in each replication. An additional 28 *S*s were discarded from the experiment. Of these losses, 4 were due to apparatus failures, 14 to *E* errors, and 10 to failure of *S* to finish the experiment within the time allotted. This last category consisted of two *S*s from each of the A–B$_{WC}$, A–B$_{BC}$, A–B$_{AC}$, and C–B$_{BC}$ conditions, and one *S* from each of the A′–B$_{BC}$ and C–B$_{AC}$ conditions.

Apparatus and Procedure

All learning was by the anticipation method at a 2:2 sec. rate with a 4-sec. intertrial interval. Words were presented on two Industrial Electronic Engineers Series 80000 digital display units mounted immediately adjacent to one another on the wall of a booth in which *S* was seated. Stimuli were programmed by paper tape which advanced every 4 sec., paced by a motor-driven cam. The stimulus term appeared on the left-hand unit as soon as the tape advanced, while the response term, appearing on the right-hand unit, was delayed for 2 sec. by a Hunter electronic interval timer. Five orders of item presentation were used for all lists. Selection of the particular orders attempted to minimize possible systematic effects of presentation order during learning on the order of output at recall. The criteria used in determining these orders were: (*a*) No item should be immediately followed by another item belonging to the same category; (*b*) no two items should follow one another more than once; (*c*) no item should appear in the same serial position more than once; and (*d*) the last item in one order should not be the first in the succeeding order. Starting order for first-list learning was determined systematically prior to the experiment, while, for second-list learning, it was the one which would have occurred next if List 1 acquisition had continued. For both lists, *S*s started to anticipate the response terms on the second trial and continued until they reached a criterion of 12/12 correct. The interval between lists was approximately 2 min.

Immediately following List 2 learning, *S*s were given a series of arithmetic problems to solve within a 2-min. period. This filled interval following learning was designed to prevent rehearsal of items and to minimize recency effects on the two free-recall tests that followed.

On the free-recall trials, *S*s were instructed to write down on a test sheet as many of the 12 adjectives as they could recall in the order in which they recalled them. The *S*s were allowed 2 min. for each recall test, with 1.5 min. between tests. The *S*s continued to work on the arithmetic problems during this interval.

Following the free-recall test, *S*s were given a sheet of paper which showed the 12 adjectives typed down one side in a random order with two blanks opposite each adjective. The *S*s were instructed to write down the two nouns with which the adjective had been paired during learning and to indicate whether that noun had appeared with the adjectives in the first list or second list.

Results

Acquisition

Table 1 presents the basic results of first- and second-list learning for the various conditions. All means are based on the combined data from the two sets of lists used under each condition as there were no significant differences between list sets.

First-list learning. Analyses of mean trials to criterion in original learning failed to show any significant differences between conditions, either by overall analysis of variance, F (7, 176) < 1, or by Duncan's test using the .05 level of significance.

The amount of grouping in first-list learning was assessed by the proportion of overt errors representing confusions among response terms paired with stimuli from the same category. On the basis of chance, these within-category errors would be expected to account for 18% of the total overt errors. As in previous studies using categorized lists, this type of error occurred with a much greater frequency than would be expected by chance in all eight conditions (Table 1).

Second-list learning. Acquisition performance on List 2 was examined for early transfer by means of number of total errors (overt and omission errors) on the first test trial, and for overall transfer by means of number of trials to criterion. Analyses of variance revealed a significant overall difference between conditions for both measures, F (8, 198) = 5.08 for errors on Trial 1, and 7.57 for trials to criterion, $p < .01$ in both cases. Specific comparisons between conditions were again made by Duncan's test at the .05 level of significance. The interesting comparisons for both measures are: (*a*) between conditions differing in type of re-pairing but identical in interlist stimulus similarity; (*b*) between conditions differing in interlist stimulus similarity but identical in type of re-pairing; and (*c*) between the transfer groups and the A–B control.

Despite the interesting pattern of results obtained when different re-pairing group means are ranked from best to poorest performance, i.e., WC > BC > AC on the first trial, and the reverse, WC < BC < AC, overall, none of the differences between types of re-pairing was significant for either measure.

The rank ordering of means for groups differing in interlist stimulus similarity remained the same for all types of re-pairing for both measures. Performance was best with stimuli representing completely different categories from the first list, second with new stimuli representing the same first-list categories, and poorest when the same specific stimulus items were used in the two lists. All differences between the A and C stimuli and four of the six differences between A and A′ stimuli were significant at the .05 level (the exceptions were the BC and AC comparisons on the first trial). Of the differences between the A′ and C stimuli, only the BC comparison on the first trial reached significance.

Comparisons of the transfer groups with the control group showed significantly superior performance for all groups with A′ and C stimuli in terms of overall performance, and for both C stimulus groups and the A′–B_{WC} group on the first trial. Although the groups with identical stimuli tended to learn more slowly than the control group, none of the differences reached statistical significance. It should be noted, however, that these results do not accurately reflect the amount and direction of specific transfer since learning to learn is also a factor in comparisons with the A–B control.

Evidence for transfer of groupings was obtained in part from an analysis of the proportion of within-category errors as defined by the second-list pairings. Since there were a large number of *S*s who made only a few or no overt errors, especially in the C stimulus groups, the error data were simply pooled over all *S*s in each condition. An overall pattern can be detected quite easily, despite the unknown reliability of the size of the differences. The two WC conditions were the only ones in which the proportion of within-category errors was sizably greater than that made by the control group in both early and overall performance. BC and AC re-pairing resulte in fewer of these errors than the contro group with identical or similar stimuli in the lists, but about the same amount when the categories of stimuli were completely different.

Free-Recall Tests

Table 2 presents the mean number of adjectives recalled on the first and second free-recall tests, the mean number of observed repetitions, $\bar{O}$(SCR), and the mean difference between obtained and expected number of repetitions, $\bar{D}$(SCR).

Number of items recalled. There was no evidence of any differences between con-

TABLE 2

MEAN NUMBER RECALLED AND MEAN CLUSTERING SCORES[a] ON FIRST AND SECOND FREE-RECALL TESTS

Measure	Condition								
	A-B	A-B$_{WC}$	A-B$_{BC}$	A-B$_{AC}$	A′-B$_{WC}$	A′-B$_{BC}$	A′-B$_{AC}$	C-B$_{BC}$	C-B$_{AC}$
Free recall I									
Mean recall	10.04	10.71	10.63	10.71	10.29	10.41	9.71	10.46	9.92
$\bar{O}$(SCR)	2.50	3.25	1.96	1.50	2.42	2.08	1.58	2.46	1.96
$\bar{D}$(SCR)	.86	1.45	.21	−.26	.73	.37	0.00	.67	.34
t	3.81**	5.51**	.89	−.96	2.33*	1.22	.02	2.12*	1.37
Free recall II									
Mean recall	10.41	10.83	10.21	10.67	10.71	10.33	10.04	10.46	10.17
$\bar{O}$(SCR)	2.08	3.13	1.92	2.08	2.63	1.96	1.92	2.96	2.08
$\bar{D}$(SCR)	.39	1.31	.25	.34	.86	.34	.28	1.21	.38
t	.80	3.62**	1.03	1.46	2.66*	1.60	1.19	5.04**	1.30

[a] Mean clustering scores are based on Bousfield and Bousfield's (1966) deviation method: $\bar{O}$(SCR) = mean number of observed repetitions; $\bar{D}$(SCR) = mean difference between obtained and expected number of repetitions.

* $p < .025$, one-tailed test.

** $p < .005$, one-tailed test.

ditions in number recalled for either recall trial, F (8, 198) = 1.60 and 1.29 for the first and second recall trials, respectively, $p > .05$. An overall mean difference of .36 items between the two list sets did reach significance on the second free-recall trial, F (1, 198) = 4.91, $p < .05$, but list set did not interact significantly with conditions, F (8, 198) < 1.

Clustering. The second index of grouping transfer was clustering of the response terms, measured in relation to the stimulus categories used in the second list. Thus the three adjectives paired with each set of categorized stimuli defined the response categories for all conditions.

Assessment of clustering was based on Bousfield and Bousfield's (1966) deviation measure, whereby the number of repetitions expected by chance, E(SCR), is subtracted from the number actually obtained, O(SCR), to yield a difference score, D(SCR), the amount of clustering not attributable to chance.

The $\bar{O}$(SCR) and $\bar{D}$ (SCR) scores for the first and second free-recall trials for the nine groups are shown in Table 2. Analyses of the D(SCR) scores by one-tailed t tests (df = 23) showed significantly above chance clustering for the A–B$_{WC}$, A′–B$_{WC}$, and C–B$_{BC}$ groups on both trials, and for the A–B control on the first trial. None of the differences were significant for the other conditions. The t values are also presented in Table 2.

Since response groupings for the A–B$_{AC}$, A′–B$_{AC}$, and C–B$_{AC}$ conditions differed in the two learning tasks, clustering was also measured in relation to the first-list pairings. The $\bar{D}$(SCR) values were −.57, .37, and .14 for the first trial, and −.40, −.04, and .09 for the second trial for the three groups, respectively. The .37 difference for the A′–B$_{AC}$ condition was significant at the .05 level, one-tailed test.

Stimulus Term Recall

The mean number of correctly recalled nouns from the first and second lists for each group are shown in Table 3. Analyses were performed on the combined data for the two list sets used under each condition. Analysis of variance showed that List 2 noun recall was significantly superior to List 1 recall overall, F (1, 184) = 239.79, $p < .01$, and Duncan's test indicated that this difference held within each specific condition. There was no overall effect of conditions, F (7, 184) = 1.84, $p > .05$, but there was a significant interaction between list recalled and condition, F (7, 184) = 2.72, $p < .05$. Further analysis revealed no differences between conditions in List 2 noun recall, but significantly poorer recall for the A′–B$_{BC}$ and A′–B$_{AC}$ groups than for any of the other groups in List 1 recall.

TABLE 3

MEAN NOUN RECALL FOR FIRST AND SECOND PA LISTS

Recall	Condition								
	A-B	A-B_{WC}	A-B_{BC}	A-B_{AC}	A′-B_{WC}	A′-B_{BC}	A′-B_{AC}	C-B_{BC}	C-B_{AC}
List 1	—	8.83	8.38	8.12	8.38	7.29	7.04	8.95	8.33
List 2	11.33	11.25	10.88	11.21	10.17	10.25	10.50	10.79	10.67

Since the number of incorrect list identifications were approximately equal for all conditions and for both lists, these data will not be presented.

DISCUSSION

In answer to the questions to which this research was addressed, the following conclusions seem to be justified: (*a*) The relationship among grouped responses necessary for the groupings to be maintained is one based on common associations with specific categories of similar stimuli, and (*b*) there is little, if any, effect of response grouping on the acquisition of specific S–R associations.

The first of these conclusions is based primarily on the direction of grouping transfer as evidenced by the amount of List 2 within-category confusion errors and the occurrence of significant clustering in free recall. Increased grouping was found only in conditions in which the response terms were re-paired with stimuli representing the same categories as in the first list (WC re-pairing). When re-paired with different first-list stimulus categories, grouping transfer tended to be negative, regardless of whether the same responses were grouped into sets in the two lists (BC re-pairing), or different responses (AC re-pairing). This latter finding clearly eliminates the possibility that the responses, once grouped together, become directly associated with one another or in some way independent of the specific stimulus categories by which they were grouped in the first place. Had this been the case, BC re-pairing would also have been expected to produce positive grouping transfer for all three conditions of stimulus similarity.

A major aspect of the first conclusion is that the critical relationship among grouped responses involves stimulus categories rather than specific stimulus terms. This statement is primarily based on the almost identical pattern of within-category errors and response clustering for the A and A′ stimulus groups under all three re-pairing conditions. Such parallel results are difficult to explain if specific stimulus items rather than categories are necessary for grouping. Additional support comes from further analysis of the order of response recall on the first free-recall test for the A and A′ groups under the BC re-pairing conditions. Since each stimulus category was associated with two different sets of three responses in the two lists, recall of the responses via cuing from stimulus category names might not be expected to produce consistent clustering according to second-list pairings, but rather some alternation between the two lists. For this reason, an analysis was made of the frequencies with which a response associated with a stimulus from a given category in the first list was immediately preceded or followed by a response associated with any stimulus from the same category in the second list. In the A–B_{BC} condition, 132 of the 230 possible responses were of this type, as were 130 of the 228 possible responses for the A′–B_{BC} condition. Only about half of these responses (68/132) in the A–B_{BC} group had been paired with the same specific stimulus item in the first and second lists and, of course, none in the A′–B_{BC} group. Unless free recall proceeds differently for the various re-pairing conditions, this analysis provides further support that the response clustering in the WC re-pairing conditions can be attributed to cuing of the responses by recall of stimulus category names. This explanation also accounts for the response clustering shown by the C–B_{BC} condition, since recall of the responses by either first- or second-list category names, or both, would still be expected to produce the same response clusters.

The second conclusion from this study was that response grouping has little or no effect on the acquisition of specific S–R associations. It was hypothesized that if response grouping did have a facilitative effect on performance, one would expect conditions which showed evidence of increased grouping in the second task to learn this list more rapidly than conditions showing decreased grouping. The results clearly do not support this hypothesis. Despite some evidence of an advantage early in learning, re-pairing responses within the same stimulus categories did not lead to more rapid acquisition than re-pairing responses between or across stimulus categories.

It should be clear that the results of this particular study do not deny that facilitation could be found under appropriate conditions. There are two factors in this experiment which would counteract any facilitative influence provided by grouping. First, the transfer paradigms used would be expected to produce considerable interference from both forward and backward associations in learning the second list. Although the amount of interference should decrease as interlist similarity decreased, even this reduced amount might be sufficient to overshadow any facilitative effects. The slight superiority of within-category over between- and across-category re-pairing when different instances of the same categories were used, as opposed to identical instances, provides some support for this explanation. Second, the failure to find facilitation might be attributed to the particular similarily relationship among the stimulus terms used to form the basis for grouping. As Underwood (1964) has pointed out with conceptually similar nouns as stimuli, it could be the case that "the gross mediating responses (category names) were so strong that they interfered with the development of more precise discriminations [p. 72]." If this were indeed the case, the necessary condition for producing facilitation might be the presence of clearly defined rules for differentiating among pairs within subsets of similar stimuli and grouped responses. There is some evidence that this is the case (Runquist, 1968).

In the absence of stronger evidence to the contrary, however, there is still little basis for concluding that grouping of stimuli and responses into subsets results in faster acquisition of specific associations.

REFERENCES

BATTIG, W. F. Evidence for coding processes in "rote" paired-associate learning. *Journal of Verbal Learning and Verbal Behavior*, 1966, **5**, 177–181.

BATTIG, W. F. Paired-associate learning. In T. R. Dixon & D. L. Horton (Eds.), *Verbal behavior and general behavior theory*. Englewood Cliffs, N. J.: Prentice-Hall, 1968.

BATTIG, W. F., & MONTAGUE, W. E. Category norms for verbal items in 56 categories. *Journal of Experimental Psychology Monograph*, 1969, **80**(3, Pt. 2).

BOUSFIELD, A. K., & BOUSFIELD, W. A. Measurement of clustering and of sequential constancies in repeated free recall. *Psychological Reports*, 1966, **19**, 935–941.

JOINSON, P. A., & RUNQUIST, W. N. Effects of intralist stimulus similarity and degree of learning on forgetting. *Journal of Verbal Learning and Verbal Behavior*, 1968, **7**, 554–559.

RUNQUIST, P. A. Clustering in free recall following paired-associate learning. *Journal of Experimental Psychology*, 1970, **83**, 260–265.

RUNQUIST, W. N. Intralist interference as a function of list length and inter-stimulus similarity. *Journal of Verbal Learning and Verbal Behavior*, 1966, **5**, 7–13.

RUNQUIST, W. N. Reversal vs. nonreversal re-pairing in categorized paired-associate lists. *Canadian Journal of Psychology*, 1968, **22**, 285–293.

UNDERWOOD, B. J. The representativeness of rote verbal learning. In A. W. Melton (Ed.), *Categories of human memory*. New York: Academic Press, 1964.

UNDERWOOD, B. J., EKSTRAND, B. R., & KEPPEL, G. An analysis of intralist similarity in verbal learning with experiments on conceptual similarity. *Journal of Verbal Learning and Verbal Behavior*, 1965, **4**, 447–462.

ZAVORTINK, B., & KEPPEL, G. Designs for studying retroaction and transfer simultaneously. *Journal of Verbal Learning and Verbal Behavior*, 1968, **7**, 700–702.

(Received June 15, 1972)

Journal of Experimental Psychology
1973, Vol. 98, No. 1, 160–163

AN EXAMINATION OF THE ADAPTATION-LEVEL THEORY ACCOUNT OF TIME ERRORS[1]

HADYN D. ELLIS[2]

University of Aberdeen, Aberdeen, Scotland

An evaluation was made of Helson's 1964 adaptation-level theory applied to time-error phenomena. Thirty *S*s were assigned to three groups. The *S*s experienced standard and variable levels centered on either 50, 65, or 80 db. Three interstimulus intervals (1, 4, and 7 sec.) were employed. In accordance with the theory, a significant effect due to the prior levels was found. The different time intervals, however, produced similar downward trends for all groups, whereas Helson predicted that they should diverge in opposite directions. Thus, it was concluded that the adaptation-level theory could not account for the strictly temporal aspects of time error.

Time errors (TE) may be defined as systematic errors in the comparison of two stimuli presented sequentially. They can be positive (i.e., the first stimulus appearing greater, louder, etc. than the second) or they can be negative (i.e., the first seeming lesser, softer, etc. than the second). Positive TE has often been reported with an interstimulus interval (ISI) of less than 3 sec. (Köhler, 1923), and also when the standard and variable stimuli are fairly small, weak, etc. (e.g., Needham, 1935; Woodrow, 1933). In many cases, however, a negative TE is found that increases as the ISI between standard and variable is lengthened.

The various theories postulated to account for TE phenomena are, without exception, inadequate. Those, for example, that are based on the idea that the memory image of the standard fades so that the variable is compared with a degraded stimulus (e.g., Kinchla & Smyzer, 1967; Köhler, 1923) cannot accommodate the evidence of positive TE.

Other theorists have argued that the memory of the standard assimilates with the stimulus value intervening between it and the variable: Usually the interval is empty and may be assumed to have a value of zero, and so the image is correspondingly decreased (Lauenstein, 1932). This argument was tested by Pratt (1933) who found, contrary to the theory, that an intervening weak stimulus produced a *greater* diminution of apparent standard size than when no stimulus was interpolated. More recently the assumption that the standard assimilates with *any* stimulus interpolated between it and the variable has, itself, been questioned (Ellis, 1971, in press).

Other theoretical postures have been adopted, including that of McClelland (1943) and Stevens (1957), but in every case, they have been unable to account for both positive and negative TE as a function of ISI, and positive and negative TE resulting from employing different overall stimulus levels.

The most cogent attempt to do so was made by Helson (1964). He explains TE in terms of his adaptation-level theory, which states that all judgments are made relative to the prevailing adaptation level (AL). This is an internalized "standard," which shifts in the direction of each new stimulus encountered and represents a weighted geometric mean of all the stimuli in, say, a laboratory experiment. Time error, claims Helson, is the result of the memory of the standard merging with the prevailing AL. Normally this is below the actual level of the standard (ordinarily, lying at the arithmetic mean of the series), since the AL, being nearer the geometric mean, is usually less than the arithmetic

---

[1] The author would like to thank Jim Torrie and Austin Davidson for designing and building the apparatus, and Rosalind Heaton for running the experiment.

[2] Requests for reprints should be sent to Hadyn D. Ellis, Department of Psychology, University of Aberdeen, King's College, Old Aberdeen, AB9 2UB Scotland.

mean of a series. Thus, negative TE is easily accounted for: The standard assimilates with an AL that is less than it, and so a variable that objectively equals the standard will seem greater. Positive TE, occurring as a result of employing lower levels of standard and variable, can also be easily explained. Needham (1935), for example, presented four levels in random order based on standards of 30, 40, 50, and 60 db. The overall AL was calculated to be 44.25 db., which is above the 30- and 40-db. standards and below and 50- and 60-db. standards. Helson's theory, then, predicts that the first two should show positive TE and the latter two negative. This is precisely what Needham had observed. In addition, he noted that these tendencies became more exaggerated as the ISI was increased.

Adaptation-level theory, however, contains no provision for dealing with temporal variables (Ellis, 1971). This omission is qualified somewhat when Helson (1964) discusses TE phenomena. He then relies on Needham's (1935) own explanation of the fact that lengthening the ISI increases the propensity for positive and negative TE. Needham conjectured that, as the interval is lengthened, so the effects of preceding stimuli are enhanced. In adaptation-level theory terms, as the ISI is lengthened the comparison will be made less on the basis of the intensity of the standard and more on the prevailing AL.

If the prevailing AL is altered by previously exposing *S*s to different ranges of stimuli (called "residual contexts"), then Helson's (1964) prediction would be similar to that illustrated in Figure 1. This shows three curves based on points of subjective equality (PSE) values obtained at three ISIs. In Helson's view the PSE is identical with the prevailing AL. Thus, when the level of prior exposure is higher than the present one, this should have the effect of enhancing the memory of the standard; and, when a lower level is experienced, the AL will be below the standard, and therefore, a variable less than the standard will suffice to produce apparent equality.

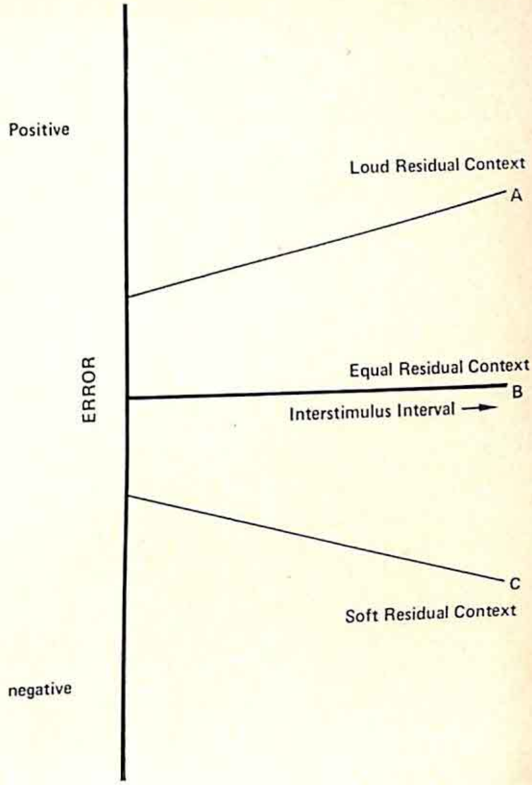


FIG. 1. Hypothetical time-error curves when standard = 65 db. for three residual context conditions based on Helson's adaptation-level theory.

According to Helson, these effects will be more pronounced as the ISI is increased.

The following experiment is an attempt to evaluate Helson's (1964) predictions concerning both the effects of residual contexts and whether any such predisposing influences increase as the ISI becomes longer.

METHOD

Subjects. Fifteen male and 15 female first-year psychology students were paid to participate. They were naive concerning both the theory and the objects of the experiment.

Apparatus. The apparatus comprised a sound source (Grason-Stadler Békésy audiometer) producing 1000-Hz. signals through an electronic switch that prevented clicks. The durations of these tones, as well as interstimulus intervals and intertrial intervals, were controlled by a tape reader (Trend). Attenuation was provided by a resistor system which allowed the variable to be ± 1 db. or equal to the standard. Up to six *S*s at a time sat in a console room. The tones appeared via earphones (Brown, Super K), and each trial was heralded by a flash from a red light. The ambient noise level in the room was a steady 38 db. (SPL).

TABLE 1

MEAN POINTS AND STANDARD DEVIATIONS[a] (IN PARENTHESES) OF SUBJECTIVE EQUALITY (IN DB) FOR THE THREE RESIDUAL CONTEXT GROUPS AT EACH OF THE THREE INTERSTIMULUS INTERVALS (ISIs) EMPLOYED

Group	ISI (in sec.)			
	1	4	7	$\bar{X}$
A	65.87 (2.05)	65.52 (.60)	65.37 (.35)	65.58
B	65.35 (.68)	64.90 (.32)	64.66 (.86)	64.97
C	65.29 (.96)	64.85 (.59)	64.40 (.97)	64.85
$\bar{X}$	65.50	65.09	64.81	65.14

[a] Between-*S* standard deviation of each mean.

Procedure. Trials consisted of a $\frac{1}{2}$-sec. warning light followed $\frac{1}{2}$ sec. later by a standard tone. An interval (1, 4, or 7 sec.) later a variable tone, ± 1 db. or equal to the standard, occurred. Both standard and variable tones lasted $\frac{1}{2}$ sec. An interval of 8 sec. then elapsed between trials.

The *S*s were semirandomly assigned to one of three groups; each group contained five men and five women. The groups received exposure to different levels in the first half of the session and then all received the same level in the second half of the session: Group A experienced 72 trials with a standard of 80 db. and variables of 79, 80, and 81 db. Group B experienced 72 trials with a standard of 65 db. and variables of 64, 65, and 66 db. Group C experienced 72 trials with a standard of 50 db. and variables of 49, 50, and 51 db.

In each case, the three ISIs (1, 4, and 7 sec.) occurred equally often and randomly. Following a 5-min. break, *S*s in each group were then subjected to a further 72 trials. This time the standard was set at 65 db. with randomized variables of 64, 65, and 66 db. for all three groups. Again the ISI was randomly 1, 4, or 7 sec. The order of trials was randomized for each subgroup tested together (1–6 *S*s at a time). Twelve practice trials with knowledge of results were given prior to the experimental trials. These comprised two examples of each variable–ISI combination at the particular intensity level initially chosen for each group.

RESULTS AND DISCUSSION

The mean PSE values for the three groups are given in Table 1. These are illustrated in Figure 2 where it is clear that the trends are not the same as the ones predicted from adaptation-level theory which are shown in Figure 1.

An analysis of variance revealed that the three groups differed in overall PSE, $F\ (2, 27) = 4.97, p < .05$. The effect of increasing the ISI was also significant, $F\ (2, 54) = 4.21, p < .05$. There is no suggestion of an interaction between groups and time intervals ($F = .04$), which would have been expected on the basis of Helson's theory.

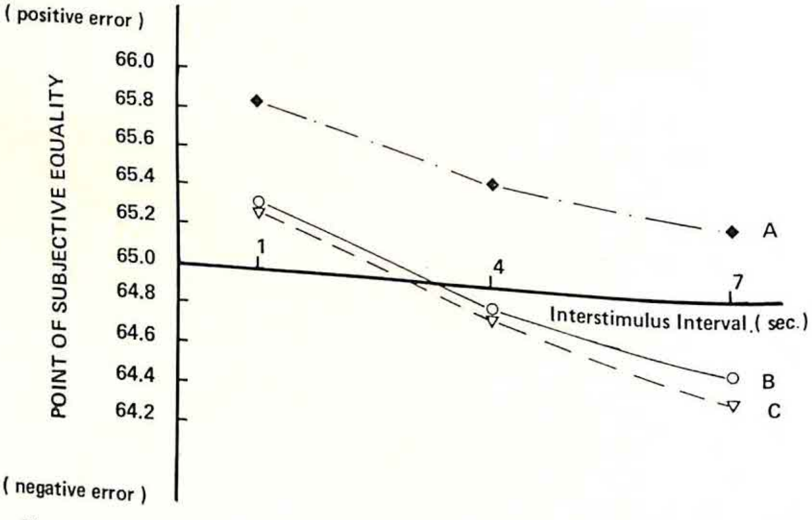


FIG. 2. The mean points of subjective equality (The individuals' points of subjective equality were derived from interpolation of their best-fitting straight line psychometric functions.) for the three residual contexts (A = 80 db., B = 65 db., and C = 50 db.), at each of the three ISIs employed (1, 4, and 7 sec.).

A posteriori analyses using the Newman-Keuls method showed that the PSEs of Group A differed from those of Groups B and C ($p < .05$), but these did not differ from each other. The 1- and 7-sec. intervals produced different PSEs for all three groups ($p < .05$), but 1- and 4-, and 4-, and 7-sec. intervals did not differ significantly.

Helson's prediction that residual context effects should affect TE has been fairly well supported. The 80-db. level context produced an overall positive TE, while the 50-db. context caused the TE to be slightly negative. The failure of the latter condition to produce a TE significantly different from that when the 65-db. context preceded is possibly due to sampling bias. If this is not the case, then Helson's theoretical approach to TE phenomena could require modification.

Certainly, his views on temporal effects do not correspond to the present data. Instead of the expected divergence in curves as the ISI increases, the three conditions were subject to the same downward trend that might be described as a tendency towards negative TE. A possible explanation of this is that, far from enhancing the effects of a residual context, increasing the time interval between standard and variable dampens it and causes some other process to occur. This might be akin to the fading-trace concept mentioned earlier, or indeed, may be the result of a change in reception sensitivity to the variable rather than any alteration in the memory image of the standard. These possibilities must await further investigation. It would appear, though, that the assertion made some time ago by Stevens (1957) that time errors are "produced by other factors than time [p. 158]" is in error.

Even when a satisfactory solution to the immediate problem is reached, there is still the occurrence of positive TE with short ISIs to explain. This is a problem that is ignored by Helson and therefore constitutes an implicit defect in his approach, which as already mentioned seems only partly appropriate.

REFERENCES

Ellis, H. D. Context effects in sensory judgment. Unpublished doctoral dissertation, University of Reading, England, 1971.

Ellis, H. D. Proactive effects of interpolated anchors. *Journal of Experimental Psychology*, 1973, in press.

Helson, H. *Adaptation-level theory*. New York: Harper & Row, 1964.

Kinchla, R. A., & Smyzer, F. A. A diffusion model of perceptual memory. *Perception & Psychophysics*, 1967, **2**, 219–229.

Köhler, W. Zur Theorie des Suksessivvergleichs und der Zeitfehler. *Psychologische Forschung*, 1923, **4**, 115–175.

Lauenstein, O. Ansatz zu einer physiologischen Theorie des Vergleichs und der Zeitfehler. *Psychologische Forschung*, 1932, **17**, 130–147.

McClelland, D. C. Factors influencing the time error of visual extent. *Journal of Experimental Psychology*, 1943, **33**, 81–95.

Needham, J. G. The effect of time intervals upon the time-error at different intensive levels. *Journal of Experimental Psychology*, 1935, **18**, 530–543.

Pratt, C. G. The time-error in psychophysical judgments. *American Journal of Psychology*, 1933, **45**, 292–297.

Stevens, S. S. On the psychophysical law. *Psychological Review*, 1957, **64**, 153–181.

Woodrow, H. Weight discrimination with a varying standard. *American Journal of Psychology*, 1933, **45**, 381–416.

(Received June 19, 1972)

Journal of Experimental Psychology
1973, Vol. 98, No. 1, 164–168

INDIVIDUAL ORGANIZATION AND RELEASE FROM PROACTIVE INTERFERENCE[1]

JOHN S. McINTYRE,[2] R. A. STOJAK,[3] AND W. MOSTOWAY

University of Manitoba, Winnipeg, Manitoba, Canada

The possibility that the encoding attributes in short-term memory (STM) and free recall are similar was investigated in the present study. *S*s were first required to sort a list of 24 unrelated words into four conceptual categories of 6 words each. Triads of words from the same sort category were then used in a Wickens release from proactive interference (PI) experiment. The main control group had four successive STM trials with items drawn from the same subjective category, while the experimental group switched to a word triad from a different subjective category on Trial 4. The experimental group scored significantly better on Trial 4, indicating that release from PI occurred and that the hypothesis of similar encoding attributes was supported.

Keppel and Underwood (1962) were able to show that much of the forgetting that occurs in a Peterson and Peterson (1959) short-term memory (STM) design may be attributable to proactive interference (PI). Keppel and Underwood found that PI builds up very rapidly over successive STM items, being near maximum after as few as three or four trials. Underwood's (1969) analysis of interference and retention processes suggests that the PI effect in such experiments is due to an overlap between the memory attributes of successive items. When memory attributes overlap, competition or interference between successive items would be expected at the time of recall, and thus, the finding of PI in STM. This argument is supported by the data of Loess and Waugh (1967) who found a decrease in PI with an increase in the intertrial interval. In addition, this reduced interference with longer intertrial intervals was partly accounted for by a decrease in the number of prior item intrusions.

Wickens (1970) has argued that if PI in STM is due to a confusion between the attributes an *S* uses in encoding successive items, then a change in one of the relevant encoding attributes should result in a release from PI. A typical release from PI experiment requires a control group which receives four successive STM trials with items drawn from a "homogeneous" pool. An experimental group receives identical items on the first three trials with the fourth item being selected for its difference along the relevant dimension. If the experimental group scores a greater percentage of items correct on Trial 4 than the control group, release from PI is said to have occurred. Wickens holds that a change in the stimulus set that leads to a release from PI indicates that this is a relevant dimension (memory attribute) along which verbal items are stored in and retrieved from memory. If no release occurs with a change in the stimulus set, it is argued that this is not a relevant encoding dimension. Using this technique, Wickens has demonstrated that maximal release is associated with a variety of semantic characteristics of STM items, and hence semantic changes are considered to be the major encoding dimensions for verbal items.

The present research is an attempt to relate encoding behavior in STM to encoding behavior in free-recall learning. That *S*s organize free-recall lists has been demonstrated repeatedly both when there are *E*-defined categorical relationships between the words of the list and when *S*s are required to impose their own organization or structure on a list of unrelated words (Tulving, 1968). Mandler (1967, 1970) has argued that *S*s must organize the words of a free-recall list into subjective units in order to recall beyond the memory span. Recall of items within the same unit is mediated by a category title or category name much as Bousfield (1953) had suggested previously. Mandler's (1967, 1970) technique for studying organization is to have *S*s sort lists of words into varying numbers of categories using any conceptual basis they wish. When a criterion of sorting consistency is reached, a free-recall test of the entire list is given. Clustering by sort category occurs at recall just as it does in the free recall of lists which have obvious categorical relationships built in. This encoding or organization in free recall must involve the use of relevant item attributes. If so, then the encoding in both STM and free recall involves encoding along item attributes, and category organization in free recall should provide information about relevant release dimensions in STM.

---

[1] This research was supported by National Research Council of Canada Grant APA-317.

[2] Requests for reprints should be sent to John S. McIntyre, Department of Psychology, University of Manitoba, Winnipeg, Manitoba, Canada.

[3] Now at John Abbott College, Montreal, Quebec, Canada.

A change in taxonomic category has been shown to be one of the most powerful STM release dimensions (Loess, 1967, 1968) as well as a primary determiner of clustering in free-recall performance (Tulving, 1968). The present research was designed to determine whether the subjective organization developed during the Mandler (1967, 1970) sort procedure would carry over into a subsequent release from PI study. More particularly, if successive STM items are from within the same sort category of a previous free-recall test, then PI should develop. If the subjective categories formed by *S*s in the Mandler procedure are parallel to taxonomic categories, then release from PI should occur when the fourth STM item is chosen from a different subjective category than the first three items.

METHOD

Subjects. A total of 132 introductory psychology students at the University of Manitoba served as *S*s to fulfill part of their course requirement. They were randomly assigned to the three treatment groups with the restriction that there be an equal number of males and females in each condition.

Learning list. The list presented for sorting and free recall consisted of 24 nouns chosen from the control condition of Stojak and McIntyre (1972). The words varied 3–14 letters in length and 3–AA in Thorndike-Lorge word frequency. No two words were high-probability free associates of one another. The words were: ANIMAL, ARMY, BLONDE, BREAKFAST, CIGARETTE, COCAINE, FEDERAL, FLOOR, ICE, MASCULINE, MONTH, MOTHER, MULTIPLICATION, NORTH, ORIENTAL, PENNY, PRINCIPAL, SPADES, SPOON, SUMMER, TENOR, VOLT, TORNADO, and YARD.

Experimental design. The design required an experimental (E) and two control groups (C_1 and C_2). Group E and Group C_1 followed the Mandler sort procedure to a criterion of two consistent sorts, and thus each *S* imposed his own organization on the list of words. Following this, Group C_1 received four successive items from within the same subjective sort category. Group E received three items from the same subjective category followed by a fourth item from a different category.

Since each STM item consisted of a three-word triad and each subjective category contained six words, it was necessary to use each of the six words of a category twice in preparing the STM items. The possibility of PI release for Group E, due to a mere change in the stimulus pool rather than a change in subjective category, was therefore present. Group C_2 controlled for this possibility by undergoing a passive sort task which did not require organizing the words, with subsequent STM items being drawn from the same pool of six words on Trials 1–3 and a new pool of six words on Trial 4.

Procedure. The *S*s in Group E and Group C_1 were given a deck of cards containing the 24 words, 1 word per card. They were instructed to sort the words into four categories containing exactly six words each. Instructions emphasized that the sort should not be done on the basis of any physical properties of the words (number of letters, alphabet) but rather on a conceptual basis. On the first sort trial, *S*s were allowed to expose all the cards before putting them in categories. Following this original sort, *S*s were required to sort the words repeatedly into four categories, being allowed to see only the top card in each category. Each word was read aloud as it was placed in a category. The cards were presented in a different random order on each trial, and sorting continued to a criterion of two identical sorts in succession. The *S*s were subsequently asked to free recall the entire list of 24 words. No time limitations were placed on the rate of sorting, but a 3-min. time limit was placed on the free-recall trial.

Group C_2 *S*s underwent a passive sort task which involved sorting the 24 cards into four categories of six words each, in much the same way as dealing a deck of playing cards. Each word was read aloud as it was placed in a category. All *S*s in Group C_2 were given five sort trials of this type with the cards in a different random order on each trial. Since there were no instructions to organize, and since items within the categories changed on each trial, organization of the words could not occur for this group. The criterion of five sorts for C_2 was based on a preliminary study which showed that almost all *S*s required to sort to a criterion of two successive identical sorts reached this criterion in fewer than five trials. Thus the actual exposure of cards for Group C_2 was as great or greater than that for either Group C_1 or Group E. Following the sort task, *S*s in Group C_2 were asked to free recall the total list of 24 words.

While *S* was writing as many of the 24 words as possible during the 3-min. recall trial, *E* constructed the eight STM items for that *S*. Each item consisted of three words from one of *S*'s four sort categories on the terminal sort trial. For Group E, Items 1, 2, and 3 were chosen from among the six words of a single category, while Item 4 was chosen from another category. Items 5–8 were chosen in a similar manner, with Items 5, 6, and 7 from the same category (but different from the categories for Items 1–4), and Item 8 was chosen from *S*'s fourth sort category. Group C_1 had four items chosen from the same sort category for Trials 1–4 and four new items from a different category for Trials 5–8. Group C_2 had items chosen identically to Group E, with the categories defined as those into which *S* sorted words on the last sort trial. Since words were put randomly into these categories and there were no instructions to organize, no conceptual relations should exist between items within a category for this group.

For all groups, no word was repeated more than twice during the four trials in which it was eligible for inclusion, and no word was ever in the same position within an item. Which categories were to be used, and in what order, was determined randomly for each *S*.

Each STM trial had the following sequence of events with cards being presented simultaneously with a click controlled by Hunter timers: Ready (3 sec.), the three-word item which *S* was required to read aloud (2 sec.), retention interval (20 sec.), and Recall (15 sec.). The Stroop test, which requires that *S* say the color in which a color word is presented, was used as a rehearsal prevention task during the retention interval. Four different Stroop cards were used, each being used with two items. A 2-min. rest was given between Trial 4 and Trial 5 to allow for PI release for all groups, since Loess and Waugh (1967) have demonstrated that PI does not build up when greater than a 2-min. intertrial interval is used. At the time the experiment was designed, the authors were not aware of the results of Kincaid and Wickens (1970) which indicate that release from PI is only 75% complete after a 2-min. rest interval following three successive STM items with a 30-sec. intertrial interval.

Two *E*s ran all *S*s individually with each *E* testing half of the *S*s in each of the conditions.

RESULTS

Of the 88 total *S*s in Group E and Group C_1 only 7 in Group E and 10 in Group C_1 failed to reach the consistency of sort criterion in the minimal three trials, and of these only 3 *S*s required more than four trials. From this it would appear that allowing *S* to see and categorize the entire list of words at once is more conducive to reaching a rapid sort criterion than the procedure used in Mandler's (1967) original experiments.

Analysis of the free-recall scores showed that Group C_2 had recall (13.2 words) inferior to that of either Group E (21.3 words) or Group C_1 (21.0 words) in spite of the fact that their five sort trials gave them more exposure to the words than the other groups. This is expected, however, on the grounds that *S* must organize the material in order to achieve high recall (Mandler, 1967, 1970).

As a check on whether *S*s in Groups E and C_1 were actually organizing the list of words, the adjusted ratio of clustering (ARC) score of Roenker, Thompson, and Brown (1971) was computed for each of the three groups with free-recall performance the dependent variable and categories defined as those into which *S* sorted on the terminal sort trial. As would be expected from the procedural differences, the mean ARC score for Group C_2 (.008) was very close to a chance score of 0 and considerably below Group E (.791) and Group C_1 (.736) which approach the maximal ARC value of 1.00 and do not differ significantly from one another. Inspection of the sort categories for all *S*s in Group C_1 and Group E suggests that certain word pairs appear together beyond a chance level (e.g., ARMY–FEDERAL, BLONDE–MOTHER, TENOR–MASCULINE, CIGARETTE–COCAINE, ICE–NORTH, TORNADO–SUMMER, and BREAKFAST–SPOON). However, beyond these diads, organization appeared to be highly idiosyncratic in nature. Postexperimental questioning indicated that most *S*s are able to place labels on their categories, but even when two *S*s reported identical or very similar category labels, the words within that category were seldom identical.

Figure 1 summarizes the recall data for the eight STM trials. Following Wickens (1970), the recall on each trial was scored out of a maximum value of four, one point for each word correctly recalled and one point for all words in the proper order.

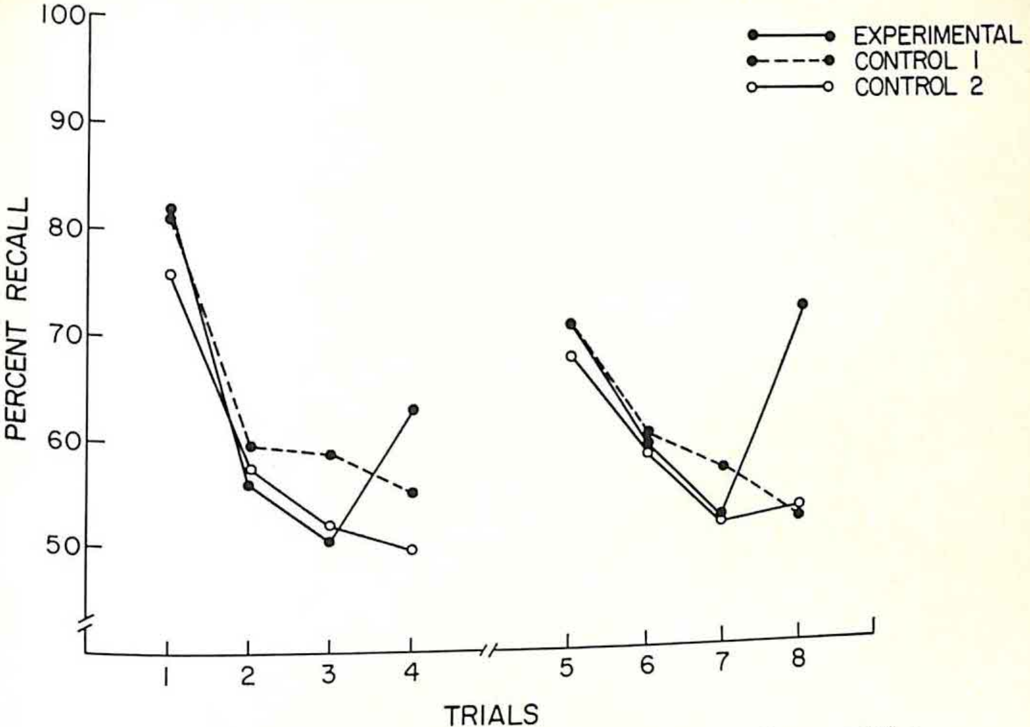


FIG. 1. Percentage of correct recall as a function of trials and test condition.

The figure shows that the expected release from PI occurred for Group E on both Trial 4 and Trial 8, with the release greatest on the second occasion. The recall scores were subjected to an arc sine transformation and an analysis of variance was performed. While it appears that the release from PI is greater for Group E on the second repetition, there was not a significant Groups × Repetition × Trials interaction, F (6, 378) = .197. The overall Groups × Trials interaction was, however, significant, F (16, 378) = 3.02, $p < .01$, with a post hoc Newman-Keuls test, indicating that (*a*) the three groups do not differ from one another on any of the first three trials of the combined sets; (*b*) only Group E shows a significant change in performance from the third to the fourth trial of a set; and (*c*) Group E performs beyond the level of either control group on the critical fourth trial. Utilizing Wickens' (1970, p. 6) procedure for calculating the percentage of release from PI, it was found that the percentages were 48.8 and 121.4 on Trial 4 and Trial 8, respectively. Groups C_1 and C_2 were combined in determining these percentage release scores.

DISCUSSION

The results of this study extend the finding of release from PI when taxonomic category is changed (Loess, 1967, 1968) to situations where categories are formed by *S*s imposing their own organization or structure on a list of unrelated words. This study both gives support to the Mandler (1967) procedure for studying organization in memory, and supports the idea that a category title or label is the basic unit of storage for free recall. The fact that no PI release occurred for Group C_2 in this experiment shows that release for the experimental group was not due to a switch to new words per se, but a switch to words that had a different representation in memory than the prior sets of items.

The results also give credence to our hypothesis that release from PI and clustering are closely associated phenomena. Where release from PI occurs with a dimensional change in the verbal materials, it should also be possible to show that manipulating this dimension will determine which items are placed in the same sort categories in the Mandler (1967) procedure and which items will cluster together on a free-recall test. We know that taxonomic category, when the categories are clearly defined, and sort category, when no general categories are present, will effect both

PI release and free-recall clustering. In addition, it has been shown that a change in grammatical class does not produce PI release (Wickens, 1970), nor do *S*s use grammatical class as a basis for clustering in a free-recall situation (Cofer & Bruce, 1965). If our argument is correct, then word frequency, semantic differential polarity, sense impression, and any other semantic attribute of words associated with PI release should also be relevant encoding attributes for free recall, and manipulating these attributes should be a predictor of which words will cluster together on a free-recall test.

Our finding that PI release can occur based on *S*-defined categories within a list of words warrants further investigation. Since our *S*s began the STM task immediately following the sort task, the question of the stability of sort categories and their relationship to PI release when a period of time separates the tasks is in doubt. Since there were no clearly defined relationships between the words of our lists, the individual's sort categories would most probably show considerable instability over time. This should lead to a reduction in the release phenomenon as the sort–STM interval increases.

REFERENCES

Bousfield, W. A. The occurrence of clustering in recall of randomly arranged associates. *Journal of General Psychology*, 1953, **49**, 229–240.

Cofer, C. N., & Bruce, D. R. Form-class as the basis for clustering in the recall of nonassociated words. *Journal of Verbal Learning and Verbal Behavior*, 1965, **4**, 386–389.

Keppel, G., & Underwood, B. J. Proactive inhibition in short-term retention of single items. *Journal of Verbal Learning and Verbal Behavior*, 1962, **1**, 153–161.

Kincaid, J. P., & Wickens, D. D. Temporal gradient of release from proactive inhibition. *Journal of Experimental Psychology*, 1970, **86**, 313–316.

Loess, H. Short-term memory, word class, and sequence of items. *Journal of Experimental Psychology*, 1967, **74**, 556–561.

Loess, H. Short-term memory and item similarity. *Journal of Verbal Learning and Verbal Behavior*, 1968, **7**, 87–92.

Loess, H., & Waugh, N. C. Proactive inhibition in short-term memory. *Journal of Verbal Learning and Verbal Behavior*, 1967, **6**, 455–460.

Mandler, G. Organization and memory. In K. W. Spence & J. T. Spence (Eds.), *The psychology of learning and motivation: Advances in research and theory*. New York: Academic Press, 1967.

Mandler, G. Words, lists and categories: An experimental view of organized memory. In J. L. Cowan (Ed.), *Studies in thought and language*. Tucson: University of Arizona Press, 1970.

Peterson, L. R., & Peterson, M. J. Short-term retention of individual verbal items. *Journal of Experimental Psychology*, 1959, **58**, 193–198.

Roenker, D., Thompson, C. P., & Brown, S. C. Comparison of measures for the estimation of clustering in free recall. *Psychological Bulletin*, 1971, **76**, 45–48.

Stojak, R. A., & McIntyre, J. S. The organization of chunks in free recall. Paper presented at the meeting of the Canadian Psychological Association, Montreal, June 1972.

Tulving, E. Theoretical issues in free recall. In T. R. Dixon & D. L. Horton (Eds.), *Verbal behavior and general behavior theory*. Englewood Cliffs, N. J.: Prentice-Hall, 1968.

Underwood, B. J. Attributes of memory. *Psychological Review*, 1969, **76**, 559–573.

Wickens, D. D. Encoding categories of words: An empirical approach to meaning. *Psychological Review*, 1970, **77**, 1–15.

(Received June 19, 1972)

Journal of Experimental Psychology
1973, Vol. 98, No. 1, 169–174

REHEARSAL AND ORGANIZATION IN INTENTIONAL FORGETTING[1]

AMOS SPECTOR, KENNETH R. LAUGHERY,[2] AND DAVID G. FINKELMAN

State University of New York at Buffalo

Two experiments explored differential rehearsal as a factor in intentional forgetting. In the first study, sequences of paired associates (CVC–word) were followed by an Only cue signaling a test from half of the list or a Both cue signaling a test from the entire list. A 3-sec. interval followed the cue and was either available for rehearsal or filled with an arithmetic problem. Performance was higher in the Only and rehearsal conditions. A Cue × Rehearsal interaction showed rehearsal had a greater effect in the Only condition. In the second study, the cue either preceded or followed the rehearsal period allowing for differential rehearsal in one condition but not the other. The effects on performance were similar. Intrusion error analyses in both studies showed a cue effect on the pattern of intrusions but no effect of rehearsal. The effects on performance are attributed to both rehearsal and organization, while the intrusion results are attributed only to organizational factors.

When *S*s are presented with a list of items to memorize and, before testing, are instructed to forget part of the list (or to remember part of it), their performance on the to-be-remembered (TBR) items improves as compared to performance on similar conditions without the forget instruction. This intentional forgetting effect has received considerable attention recently, and the result holds even if the forget instruction is given after presentation of all items (postinput instruction). One possible explanation of this phenomenon proposed by Bjork (1970) states that after the instruction to forget, *S* can differentially rehearse only the TBR items, thereby increasing later recall.

The differential rehearsal interpretation has been investigated by Shebilske, Wilder, and Epstein (1971). They presented *S*s with lists of four paired associates (CVC–word) with 3-sec. intervals after the second and fourth pairs. These intervals were either both filled with an arithmetic problim (filled condition), or only the second interval was filled (semifilled condition), or neither was filled (unfilled condition). Before the second interval, a cue was presented indicating from which set of pairs the subsequent test would come. The cue signaled either the first two pairs, the second two pairs, or the entire set. Another variable manipulated in the experiment was categorized vs. uncategorized lists. In the categorized condition the words in any one list were all members of a definable class. In comparing the semifilled with the unfilled conditions, they found that the cue increased performance by the same amount in both conditions. This result contradicts the interpretation of postinput differential rehearsal, since it was assumed that in the semifilled condition *S* was busy solving the arithmetic problem and could not rehearse. If this assumption is true, and if *S* rehearses the TBR set in the unfilled condition, an interaction should result: The cue should help more in the unfilled than in the semifilled condition. Shebilske et al. argued that differential rehearsal does not contribute significantly to the cue effect. They proposed instead an organizational mechanism that organizes the input by sets, thereby restricting the search set for the conditions where the cue signals only half of the list.

The results obtained by Shebilske et al. (1971) are inconsistent with current theoretical formulations. To begin with, differential rehearsal and organization are not exclusive. Since the cue comes after the input, differential rehearsal presupposes some kind of organization because *S* must be able to identify in memory the TBR set in order to rehearse it. In addition, the importance of rehearsal has been shown in innumerable memory experiments, and rehearsal processes are an important part of many current memory models. Shebilske et al. did not get a rehearsal main effect. Why in this experiment should an opportunity to rehearse items not enhance the recall of those items? The results might be an artifact of some peculiarity of the experiment. The performance levels in this experiment were high, especially in a condition using categorized lists (70%–90%), which might result in a ceiling effect. This possibility, however, cannot explain the same results obtained in another study by Epstein, Massaro, and Wilder (1972), where performance was much lower. Another possibility is that the interpolated task was too easy and allowed some rehearsal even in the filled and semifilled conditions. It should be noted, though, that *S*s' introspective reports indicated the task was sufficient to prevent rehearsal.

---

[1] This research was supported by Grant MH-11595 from the National Institute of Mental Health, U. S. Public Health Service.

[2] Requests for reprints should be sent to Kenneth Laughery, who is now at the Department of Psychology, University of Houston, Houston, Texas 77004.

The following experiments were designed to explore these possibilities. The first experiment was essentially a replication of part of the Shebilske et al. (1971) experiment. The study dealt with the issue of task difficulty by using two different list lengths. It was expected that if a ceiling effect or an insufficient interpolated task is the reason for their results, increasing the length of the list will result in a triple-order interaction. The Cue × Rehearasl interaction (the differential rehearsal effect) will exist in the longer lists, and not in the short lists. Experiment II was designed as a replication of part of Experiment I with a different manipulation of rehearsal opportunity.

METHOD

Experiment I

Subjects. Sixty-four *S*s (34 males and 30 females) participated in the experiment. Participation served as partial fulfillment of an introductory psychology course requirement. The *S*s were assigned randomly to the experimental conditions.

Apparatus and materials. The stimulus materials consisted of CVC–word pairs. The CVCs were taken from the Archer (1960) norms, ranging between 37–62. No CVC involving a Y was used. The words were high-frequency three- or four-letter words. CVC–word pairing was done randomly. Two sets of materials were composed, each containing 32 lists. One set had four pairs in each list, while the other had eight. The presentation of each of these lists was followed by a blank interval (rehearsal [R] condition) or by an arithmetic problem (no rehearsal [NR] condition). The arithmetic problem was always a two-digit number minus a one-digit number, such that a 10 boundary would be crossed for a correct answer (e.g., 38 – 7 would not be allowed; 38 – 9 would be). The lists were presented visually, using a slide projector.

Procedure. The two list lengths and two rehearsal conditions resulted in four treatment combinations with 16 different *S*s run in each. The sequence of events for each list was: two (or four) pairs for 2 sec. each, a 3-sec. blank interval (always blank), another two (or four) pairs for 2 sec. each, a 300-msec. cue, and a 3-sec. interval that was either blank or filled with an arithmetic problem. The arithmetic problem was presented visually. After the last 3-sec. interval, a CVC from the list was shown and *S* had to write the appropriate word response. The CVC test item stayed in view for 8 sec. and was followed by a 1-sec. ready slide (blank red) and then the next list.

The 300-msec. cue consisted of two lights, one to the left and one to the right of the viewing screen. The left light signaled that the test would be from the first two (or four) pairs, the right light indicated a test from the last two (or four) pairs. The occurrence of both lights signaled that the test could be from any item in the list. Both lights (both condition) were turned on in 16 lists, and the left light and right light (Only condition) in 8 lists each. Each serial position was tested an equal number of times in each of the three cue conditions. These different cuing conditions were assigned to the 32 lists in a semirandom way: The 32 lists were divided into three blocks containing 12, 12, and 8 trials. In each block, half of the lists had a Both cue and one fourth had one of the Only cues. These signals were assigned randomly within each block. This procedure was used to prevent long runs of a particular cue; the longest run was three Both cues in a row and appeared only once in the experiment.

Two practice lists preceded the presentation of the 32 lists. The cues in the practice lists were left and right, in that order.

The *S*s were instructed about the cue and were reminded again about it after the two practice lists. They were also told (for the arithmetic condition) to write the answer to the arithmetic problem before responding to the test CVC. Guessing was encouraged. The *S*s were run in groups, varying in size between 7 and 10.

Experiment II

In order to increase the generality of the first experiment, a second study was carried out using different *S*s and different stimuli. The major change, however, was_the manipulation of rehearsal opportunity. In Experiment II the interval was always blank and the timing of the cue was varied. In one condition the cue came before the 3-sec. interval, and in the other it appeared after the interval, just before the test stimulus In the first condition *S* had an opportunity to rehearse differentially the TBR set. In the second condition he did not, since during the rehearsal period he did not know which set to rehearse. The conditions were designated DR (differential rehearsal) and NDR (no differential rehearsal). The prediction is that differential rehearsal contributes to the intentional forgetting effect, and a Cue × Rehearsal condition interaction will result.

Subjects. Forty *S*s (14 males and 26 females) participated in the experiment. Participation served as partial fulfillment of an introductory psychology course requirement. The *S*s were assigned randomly to the experimental conditions.

Apparatus and materials. Only four-pair lists were used in this experiment. The lists were formed in the same way as those in Experiment I but with different pairs.

Procedure. The procedure was the same as in Experiment I with the necessary changes in instructions to fit the new condition. The *S*s were run in groups of 4–8.

RESULTS

The results of these two experiments are presented in order of the two methods used to analyze the data (performance and intrusion analyses) rather than in order of the two experiments. The reason for this departure from the usual format is that the findings make this presentation more convenient and meaningful.

Performance Analyses

Table 1 shows performance in terms of proportion of correct responses for Experiments I and II.

Experiment I. An analysis of variance was carried out with two between-*S* variables (list length and rehearsal condition) and one within-*S* variable (cue). All main effects were significant. For list length, F (1, 60) = 69.35, $p < .01$; for cue, F (1, 60) = 37.57, $p < .01$; and for rehearsal, F (1, 60) = 23.60, $p < .01$. Thus, in addition to the usual list length effect, both the rehearsal effect and the intentional forgetting effect were well confirmed.

The critical Cue × Rehearsal interaction was also significant, F (1, 60) = 5.21, $p < .05$. The data indicate that the cue effect was stronger in the rehearsal condition than in the nonrehearsal condition. It should be noted, however, that the cue effect does exist in the nonrehearsal condition. A separate analysis showed this effect to be significant, F (1, 30) = 7.69, $p < .05$.

TABLE 1

PROPORTION OF CORRECT RESPONSES

Cue condition	Experiment I				Experiment II	
	Four-item list		Eight-item list			
	R	NR	R	NR	DR	NDR
Only	.68	.39	.30	.21	.60	.50
Both	.50	.32	.20	.16	.50	.50

Note. Abbreviations: R = rehearsal; NR = no rehearsal; DR = differential rehearsal; and NDR = no differential rehearsal.

The only other significant result was the Rehearsal Condition × List Length interaction, F (1, 60) = 7.02, $p < .05$. Rehearsal helped more for the shorter lists. This result can readily be accounted for. Since for both list lengths the rehearsal period is the same (3 sec)., the four-pair lists have twice as much rehearsal time per item as the eight-pair lists. Therefore, rehearsal helps more for the shorter lists.

Experiment II. The data from this experiment were analysed using the planned comparisons technique. The results yielded a significant cue effect, t (38) = 2.002, $p < .05$, one-tailed; and a significant Cue × Rehearsal Condition interaction, t (38) = 1.69, $p < .05$, one-tailed. The rehearsal condition did not have a significant effect, t (38) = 1.24, $p > .10$. Table 1 shows that the cue helped only when there was an opportunity for differential rehearsal.

Intrusion Analyses

The types of incorrect responses *S*s can make include omissions, words that did not appear in the list, or words that did appear in the list. The last type of error we call intrusions. Intrusions constituted 55% of the incorrect responses in Experiment I, and 63% in Experiment II.

An intrusion can come from the same sublist (set) that is tested, or from the other set. For example, if *S* is tested on

TABLE 2

Proportion of Same Set (SS) and Different Set (DS) Intrusions

Type of intrusion	Experiment I				Experiment II	
	Four-item list		Eight-item list			
	R	NR	R	NR	DR	NDR
Only condition						
SS	.33	.29	.32	.29	.44	.39
DS	.33	.28	.21	.26	.19	.25
Both condition						
SS	.20	.17	.21	.24	.22	.19
DS	.40	.41	.34	.32	.44	.37

Note. Abbreviations: R = rehearsal; NR = no rehearsal; DR = differential rehearsal; and NDR = no differential rehearsal.

the fourth item (in a four-item list) and responds with the word from the third pair, his intrusion comes from the same set as the test. If he responds with the word from the first or second item, the response is from a different set.

Each *S*'s intrusions were counted and classified into SS intrusions (same set) or DS intrusions (different set). They were further classified into intrusions that occur on Both trials and those that occur on Only trials. For each of the experimental conditions, proportion of SS and of DS intrusions out of the total number of errors in that condition were calculated. These proportions are presented in Table 2. Two-way analyses of variance were performed on these data, with cue condition as a within-*S* variable and rehearsal condition as a between-*S* variable. Separate analyses were performed on the two list lengths of Experiment I and Experiment II. The data for two *S*s in Experiment II were randomly discarded to create equal *n*s as a convenience for the particular computer program used in the analyses. In one set of analyses the proportion of SS intrusions was the dependent variable, and in the other set, the proportion of DS intrusions.

In all six analyses (three experiments and two dependent measures) only the cue effect was significant ($p < .05$ in all cases). It can be seen from Table 2 that Only conditions result in more SS intrusions and less DS intrusions than Both conditions. As mentioned above, this was the only significant effect. The rehearsal effect and the Rehearsal × Cue interaction are not significant in any of the six analyses. Ten of the 12 *F*s are less than 1 and for the highest F, $p > .10$.

A three-way analysis of variance was also carried out for each of the three experiments, using type of intrusion (SS vs. DS) as a third independent variable. It can be easily seen that the results of the previous analyses are contained in this one. They were introduced to facilitate the presentation. The three-way analyses showed that type of intrusion interacts only with the cue variable, which reflects the fact that there are more DS than SS intrusions in the Both conditions, and the opposite in the Only conditions. No other interaction involving type of intrusion approached significance.

An important thing to note in Table 2 is the relation between the proportion of SS and of DS intrusions across the treatment combinations. In the four-item lists, given that *S* makes an error from the same list, he has two opportunities for a DS intrusion and one for an SS intrusion. In the eight-item lists, the numbers are four and three, respectively. Table 2 shows a tendency to give more SS than DS intrusions in the Only conditions, and more DS than SS intrusions in the Both conditions. Furthermore, the ratio SS/DS for the Both conditions approaches 1/2 in the short lists and 3/4 in the long lists. These results do not interact with rehearsal condition.

Discussion

In Experiment I it was expected that while the results might not show a differential rehearsal effect for the four-item lists, that effect would be seen in the longer lists—a triple-order interaction. However, the differential rehearsal effect was significant regardless of list length. These results are confirmed by the performance analysis of Experiment II and stand in direct contradiction to Shebilske et al. (1971) and Epstein et al. (1972). Our intrusion data also contradict Shebilske et al. In their noncategorized lists, even in the Both conditions, there were always more SS than DS intrusions. We do not know how to account for these discrepancies. The possibility of a ceiling effect in the Shebilske et al. study exists. Our *S*s' performance (about 60% in the highest performance condition) was lower than theirs. However, in the Epstein et al. experiment a ceiling effect cannot account for the lack of interaction, since their *S*s performed at about a 45% level.

Further insight as to the role of rehearsal and of organization in the experiment is provided by the intrusion analyses. Intentional forgetting yields two effects: the improvement of performance when *S* is responsible for only a part of the list, and a decrease in intrusions from the TBF items as compared to the appropriate control. Bjork (1970) has suggested that these two effects may be accounted for by a decrease in interference. While the results of our experiments show that the cue has a strong effect on both performance and intrusion error pattern, the effect of the rehearsal variable is quite different. Rehearsal has a strong effect on performance but no effect on the pattern of intrusions. This is also reflected by the Rehearsal × Cue interaction; differential rehearsal affects performance but not intrusion error pattern.

Usually, rehearsal is defined by letting *S* have some time with no interpolated task. The assumption is made that he rehearses, repeatedly saying the items to himself subvocally. However, during this time for rehearsal other processes might also be employed, such as organizing, categorizing, creating mneumonic devices, etc. If one assumes that in these experiments *S* organizes the input according to the two sets that make up a list (by tagging the items, storing them in different locations, or any other procedure that will enable him to know to which set an item belongs), the obtained pattern of results might be expected. A cue will restrict the search set and thereby increase performance on a mere probabilistic basis. Furthermore, the pattern of intrusions when the cue is Both should be as it is: SS/DS should be 1/2 or 3/4, according to list length.

The fact that rehearsal did not interact with cue in the intrusion analysis, indicates that in this study rehearsal has no effect on the organizational process. Whatever is happening during the blank time interval in the rehearsal groups, it does not contribute to the organization of the list into two sets. Rehearsal does, on the other hand, have a strong effect on performance. It increases performance, and this increase is greater for the Only conditions.

One notion about the effect of rehearsal is to say that it increases item strength (or trace strength). If so, then differential rehearsal should be reflected in the intrusion pattern. Our results are inconsistent with this notion. Increasing item strength should give relatively more SS intrusions when it is done only on two items rather than on four (or four rather than eight). In other words, when the cue was Only and *S* failed to respond correctly, he is left with one strong item and two weak ones in the rehearsal conditions as compared to three weak items in the nonrehearsal conditions. When the cue was Both, he has either three strong items or three weak items. If he makes his choice according to item strength, an interaction should be observed. The results show this does not happen. Another finding counter to the item strength notion stems from the number of intrusions in the R and NR conditions. Percentage of intrusions of all incorrect responses should be higher if item strength increases and should, therefore, be higher for the rehearsal conditions. The observed data for the short lists in Experiment I was 55% and 59% for NR and R conditions, respectively, 55% and 53% for Experiment I long lists, and 66% and 60% for Experiment II. From these results we conclude that rehearsal does not increase item strength in this sense.

It seems possible to account for most of our data with an alternative model. Suppose that *S* organizes the items into two groups. This organizing can be done during input and is not dependent upon postinput time. When there is an opportunity to rehearse he rehearses the TBR set, which means that he rehearses differentially. This rehearsal discriminates among the items of the TBR set, without increasing their relative strength. At output, the two processes have an effect. The restriction of the search set by the cue and the better discrimination between the TBR items explain the two main performance effects. The differential rehearsal effect is explained by the fact that discrimination is needed among half the items in the list when the cue is Only, and for all list items when the cue is Both, while the amount of time available for rehearsal is the same in the two conditions. When an incorrect answer has been given, it is assumed this discrimination "failed to work" for *S* on that specific trial. As a result of the cue, however, he will choose a response from the restricted search set, thus showing the observed pattern of intrusions.

One observed result is not explained. In the nonrehearsal conditions the cue should affect performance because it restricts the search set. This effect occurs in Experiment I where performance is better in the Only conditions, even when rehearsal is prevented. But in Experiment II, where there is no opportunity for differential rehearsal, performance on Only and Both trials is almost exactly the same. Restriction of the search set did affect the intrusion pattern, but it should have been reflected in performance also. It is tempting to say that Experiment II was a different situation, but except for this deviate result, all other data are essentially the same as in Experiment I.

Whether this model is valid or not, and whether it applies to other experimental situations, is conjecture at this point. Other models could be developed to account for the results. One thing seems certain: Any model to account for these results should make a clear distinction between organization and rehearsal and between normal performance data and intrusion data.

REFERENCES

ARCHER, E J. A re-evaluation of the meaningfulness of all possible CVC trigrams. *Psychological Monographs*, 1960, **74** (10, Whole No. 497).

BJORK, R. A. Positive forgetting: The non-interference of items intentionally forgotten. *Journal of Verbal Learning and Verbal Behavior*, 1970, **9**, 255–268.

EPSTEIN, W., MASSARO, D. W., & WILDER, L. Selective search in directed forgetting. *Journal of Experimental Psychology*, 1972, **94**, 18–24.

SHEBILSKE, W., WILDER, L., & EPSTEIN, W. Forget instructions: Effect of selective rehearsal and categorical distinctiveness. *Journal of Experimental Psychology*, 1971, **89**, 372–378.

(Received June 19, 1972)

Journal of Experimental Psychology
1973, Vol. 98, No. 1, 175–183

ACCURACY OF PSYCHOPHYSICAL JUDGMENTS AND PHYSIOLOGICAL RESPONSE AMPLITUDE[1]

ROBERT J. GATCHEL AND PETER J. LANG[2]

University of Wisconsin—Madison

The effects of an auditory judgment task on physiological responding were investigated in two separate experiments. The *S*s in both studies were instructed to discriminate pure tones varying along a particular stimulus dimension. The stimulus set in Experiment I consisted of five tone intensities. In Experiment II, five tones differing in frequency were presented. In both studies, the difference between the initial acceleratory peak and the subsequent deceleration of the cardiac wave form was found to reliably differentiate between stimuli on the basis of judgmental accuracy. Cortical slow-wave activity was recorded in Experiment II and was also found to covary with accuracy of discrimination.

This study was designed to test the hypothesis that the amplitude of the stimulus-evoked heart rate wave form is systematically related to the accuracy of subsequent psychophysical judgments. The experiment was prompted by a recent study of heart rate (HR) and auditory stimulus intensity by Bull and Lang (1972). In order to insure that their *S*s attended to the stimuli, these investigators required *S*s to rank order from highest to lowest each of five intensities presented. Analysis revealed that while the evoked heart rate response increased with increments in stimulus intensity as predicted, wave form amplitude also covaried with the accuracy of stimulus judgments. In effect, those stimuli which were discriminated most readily prompted the largest HR responses; tones which were confused with each other, despite repeated exposure, evoked small HR change.

The present investigation included two separate experiments. The stimuli in Experiment I consisted of five different auditory intensities, as in Bull and Lang (1972). The subjective distances between the five tone levels were manipulated in order to make the two end points more easily discriminable than the center tones. It was predicted that the HR-evoked potential would be larger in amplitude to high-intensity than to low-intensity tones, and furthermore, that it would be relatively attenuated to center tones (which evoke more errors in psychophysical judgment) than to the well defined anchor stimuli. Furthermore, it was expected that the effects of stimulus judgment on heart rate would persist after the intensity response had habituated. Thus, the experiment represented an effort to directly cross-validate the Bull and Lang (1972) results. It differed from the latter study in that the subjective distances between stimuli were more exaggerated, so as to show the effects of judgmental response on heart rate at an earlier stage in the stimulus series. Furthermore, a reaction time (RT) task was included along with the psychophysical judgments in the Bull and Lang (1972) study; this potentially confounding influence was eliminated in the present design.

In Experiment II, *S*s were asked to discriminate auditory frequencies rather than intensities. Again, the stimulus series was manipulated so that the subjective distance between the anchor stimuli and the auditory frequency of adjacent tones was greater than it was among the three center tones. There is no evidence that frequencies in the middle of the audible range produce different evoked HR responses. Thus, it was predicted that the distribution of HR response amplitudes would be determined solely by the relative discriminability of the stimuli.

---

[1] This research was supported in part by grants to the second author from the National Institute of Mental Health (MH-10993 and MH-35,324) and the Wisconsin Alumni Research Foundation. Parts of both experiments were conducted while the first author was a National Institute of Mental Health predoctoral fellow (MH-50,779), and a portion of the data was presented by the first author in partial fulfillment of the requirements of the master's degree at the University of Wisconsin. The assistance of Michael Falconer in analyzing the data and of Kent Hayes in developing and maintaining apparatus is gratefully acknowledged.

[2] Requests for reprints should be sent to Peter J. Lang, Department of Psychology, University of Wisconsin, Madison, Wisconsin 53706.

Separate components of the cardiac rate response were examined to determine which measure best related to psychophysical judgments. In addition to heart rate, slow-wave cortical activity was measured in Experiment II. Both Lacey and Lacey (1970) and Connor and Lang (1969) have shown parallels between these two physiological phenomena, i.e., cortical negativity tends to increase just prior to the imperative stimulus in the RT situation, coincident with cardiac slowing. Pilot work from this laboratory suggests that changes in cortical voltage level in the first few seconds after stimulus onset may covary with the accuracy and reliability of psychophysical judgments.

EXPERIMENT I

Method

Subjects. The sample consisted of 10 male undergraduate volunteers from the introductory psychology *S* pool at the University of Wisconsin.

Apparatus. Heart rate was continuously recorded throughout the experiment. Miniature Beckman silver–silver chloride electrodes were attached to *S*'s skin over the right and left anterolateral lower ribs, and an ear clip was fixed to the left ear lobe to serve as ground. The HR signal was monitored by a Grass 5P4 EKG preamplifier. The output from this preamplifier was recorded on one channel of a seven-channel Technical Measurement Corporation Model 700/1400 series magnetic tape-recording system. Five of the remaining channels, one for each of the five stimulus tones, recorded a series of pulses, each preceding that particular stimulus onset by 2.03 sec. These pulses served as tags for later retrieval of the EKG data.

The auditory stimulus was a 6-sec. 700-Hz. tone generated by a Hewlett Packard 200 CD oscillator. The tone was amplified by a Scott Solid-State Stereomaster 299-F amplifier with a 10-msec. rise time determined by a ramp-control HS-17 electronic switch. A celluloid tape, used to control ITI (intertrial interval) lengths, triggered a Space Mechanisms photocell device at preselected intervals. The photocell, in turn, activated an electronic digital timing generator which transmitted a pulse to one of the five tape-recorder channels 2.03 sec. before stimulus onset. This timing generator then closed the electronic switch for the 6-sec. stimulus presentation. A five-position attenuator-channel selector was used to attenuate the tone to one of five intensity levels—50, 70, 75, 80, or 100 db. as measured by Scale C of a General Radio sound level meter (Type No. 15516). The channel selector also determined which tape channel would receive the pulse from the timing generator. The *E* changed the selector during the experiment according to a semirandom stimulus presentation sequence.

The tones were transmitted through a twin pair of Jensen speakers, one positioned on each side of *S*, approximately 12 in. from each ear. Judgment responses were recorded on one channel of the polygraph from a five-switch pianolike keyboard mounted on *S*'s dominant-hand chair arm. Each key selected a different voltage which affected the pulse heights when amplified and recorded on the polygraph.

Data reduction. The analogue EKG signals contained on FM magnetic tape were played into a LINC-8 Digital Equipment Corporation computer programmed to measure the time between successive R–R (Response–Response) intervals in milliseconds. These data were later transferred to digital computer tape for submission to a UNIVAC 1108 computer. The interpulse interval data were subsequently converted to a rate per second format, each beat being weighted by the proportion of a second that it was present.

Response measurement. Evoked HR response curves were obtained by averaging one prestimulus and nine poststimulus seconds over 10 presentations of each tone intensity within each series. Figure 1 is an idealized representation of the obtained cardiac response wave form. Further analyses were made based upon the following points of interest taken from these curves:

B = base line sample (average rate for 2.03-sec. prestimulus).
D_1 = slowest second during the first 2 sec. after stimulus onset.
A = fastest second subsequent to D_1 and within the first 5 sec. after stimulus onset.
D_2 = subsequent slowest second before the eighth second after stimulus onset.

Nine different component and combination of component measures were analyzed. These measures were developed by Bull and Lang (1972) and represent deviations from the base interval or difference scores between the peaks and valleys of the wave form. Of special interest was the $A - D_2$ component limb, first described by Lang and Hnatiow (1962) and shown to be an index of habituation. Also of interest were the $B - D_1$ and $B - D_2$ measures, reflecting a cardiac deceleratory trend which Graham and Clifton (1966) argue is associated with attention.

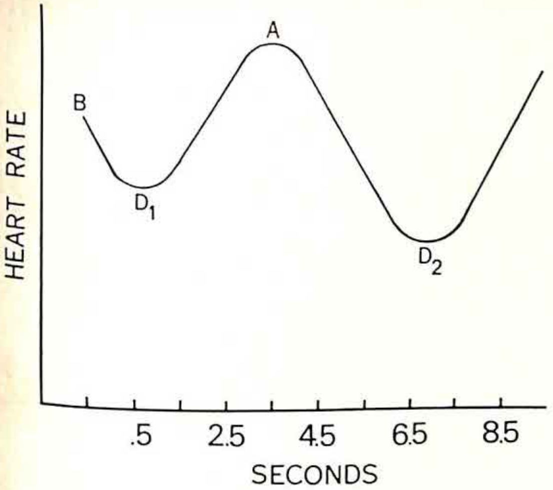


FIG. 1. Idealized representation of the evoked cardiac wave form.

Procedure. The *S*s were seated in a comfortable chair in a partially soundproof room. The experiment consisted of three stimulus series, each lasting approximately 20 min., with a 5-min. rest period between series. Each series consisted of 50 tone presentations, 10 at each intensity level. A semirandom order of presentation was employed with a mean ITI of 20 sec. and a range of 15–25 sec.

The *S* was informed that a series of tones would be presented, one at a time, and that they would vary across five intensity levels. He was told to judge the intensity rank of each tone from 1 (lowest) to 5 (highest). He was also told that the very first tone presented would be the lowest intensity. In making his judgment, *S* pressed one of the five ordered response keys following tone offset. The *S*s were not encouraged to respond rapidly, but simply to press the key whenever they made their judgment.

Results

Judgmental errors. Figure 2 shows the distribution of judgmental errors across intensities for all trials. Trend analysis indicated no basic change in the form of this curve over series, although a significant reduction in total errors was indicated: series linear component, $F\ (1, 9) = 40.83$, $p < .001$.[3] The curve shape indicates success in manipulating the discriminability of intensities. The three middle stimuli occasioned the most judgmental errors; the fewest errors were made to the extreme stimuli. This was confirmed by a large quadratic trend, $F\ (1, 9) = 74.79, p < .001$, accounting for 67% of the variance. There was also an unexpected significant quartic trend, $F\ (1, 9) = 22.79$, $p < .001$. This latter component accounted for only 24% of the variance, but it indicates a tendency for the middle intensity to be judged correctly somewhat more often than were Stimuli 2 and 4.

Three orthogonal planned contrasts of means were performed on these judgmental errors. The null hypotheses of these three contrasts were:

Contrast 1—H_0: The average of Stimuli 2, 3, and 4 = the average of Stimuli 1 and 5.

Contrast 2—H_0: Stimulus 1 = Stimulus 5.

Contrast 3—H_0: The average of Stimuli 2 and 4 = Stimulus 3.

Contrast 1 was significant, $F\ (1, 36) = 72.03$, $p < .001$, substantiating the trend of the anchor stimuli being judged correctly most often. Contrast 3 was also significant, $F\ (1, 36) = 14.92$, $p < .001$, verifying that the middle stimulus was judged correctly more often than Stimuli 2 and 4. There was no difference between Stimuli 1 and 5.

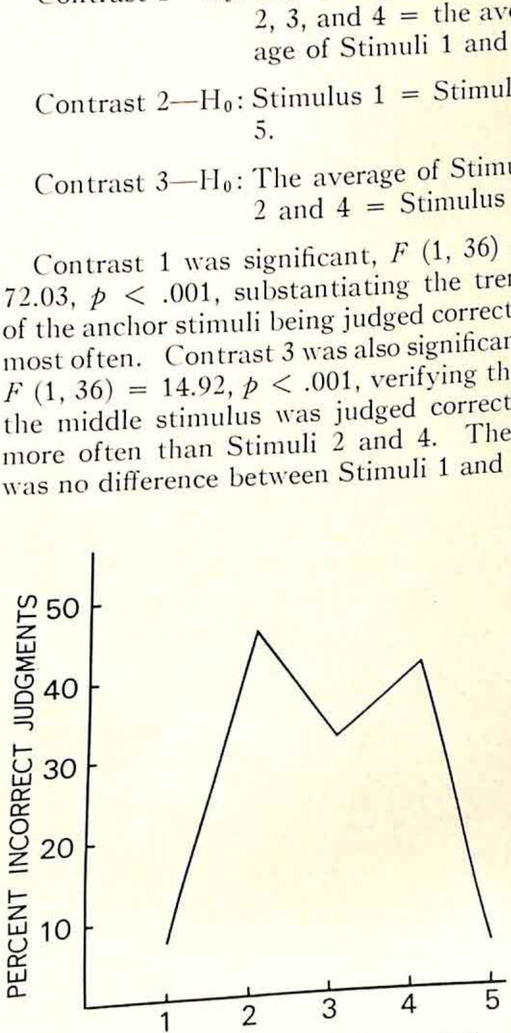


FIG. 2. Percentage of incorrect judgments for each stimulus intensity level across all series (Stimulus 5 is the loudest tone).

---

[3] The authors are indebted to Frances K. Graham (University of Wisconsin—Madison) for the use of her computer program for trend analysis.

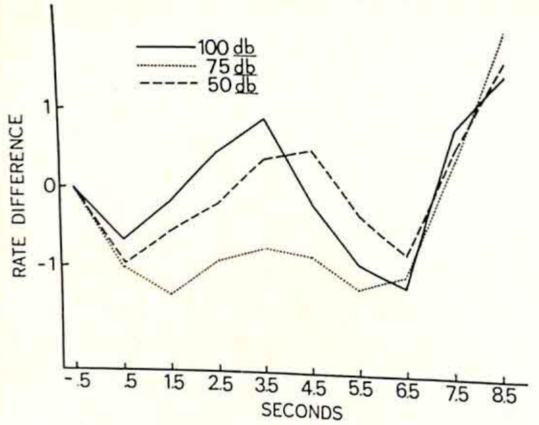


FIG. 3. Second-by-second averages of heart rate response across *S*s for the highest, lowest, and middle intensity tones for Series III—rate difference from Second −.5.

Heart rate. Averages of the HR response across *S*s for the loudest, the softest, and the middle tone, for Series III, are shown in Figure 3. To facilitate comparison, the curves have been brought to a common level at .5 sec. prior to stimulus onset.

A trend analysis of the data for all tones over all series indicated significant linear through quartic trends across seconds, F_1 (1, 9) = 16.70, F_2 (1, 9) = 11.92, F_3 (1, 9) = 17.58, F_4 (1, 9) = 46.80, all $ps < .01$. These trends confirm the triphasic character of the cardiac response, which is shown graphically in the Series III data of Figure 3. Each component was a significant aspect of the wave form (initial deceleration [D_1], acceleration [A], and secondary deceleration [D_2]).

Hypotheses were tested by analyzing the wave form component measures presented in Table 1. As for the judgment error data, three orthogonal planned contrasts of means were performed. Contrast 1 (Intensities 1 and 5 vs. 2, 3, and 4) was significant for the difference between heart rate acceleration and secondary deceleration, $A - D_2$, F (1, 36) = 5.907, $p < .025$; the initial acceleration from the prestimulus base, $A - B$, F (1, 36) = 6.347, $p < .025$; and the combination of these two measures, $(A - D_2) + (A - B)$, F (1, 36) = 6.180, $p < .025$. Thus, the main hypothesis was strongly supported: The two anchor stimuli clearly elicited larger responses than the middle tones.

Stimulus 5 was significantly larger than Stimulus 1 (Contrast 2) for the $(A - D_2) + (A - B)$ measure, F (1, 36) = 4.136, $p < .05$, and nearly significant for the other two measures. These Contrast 2 results reflect the intensity factor and are attributable to the relatively large acceleratory responses evoked by the 100-db. tone. The test of differences between the middle tones was not significant.

Analysis of the raw heart rate dataa lso indicated significant linear, F (2, 18) = 3.91, $p < .05$, and quadratic, F (2, 18)

TABLE 1

MEAN HEART RATE (HR) COMPONENT MEASURES FOR EACH STIMULUS INTENSITY LEVEL OVER ALL SERIES IN EXPERIMENT I

Stimulus intensity level (in db.)	HR wave form component measures (beats/minute)		
	$A - D_2$	$A - B$	$(A - D_2) + (A - B)$
50	2.50	.75	3.25
70	2.00	.40	2.25
75	1.60	.55	2.10
80	2.20	.60	3.25
100	4.30	1.40	5.75

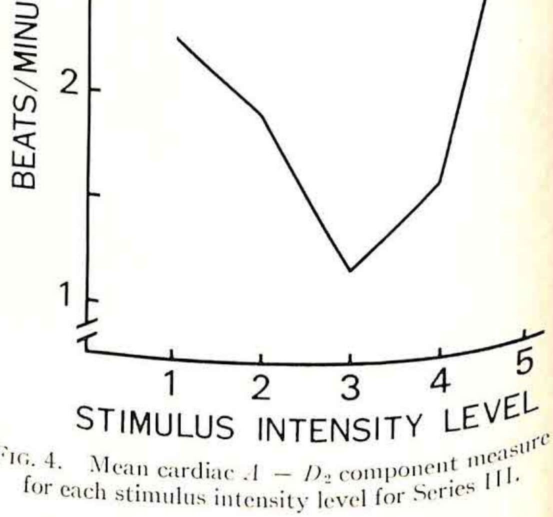


FIG. 4. Mean cardiac $A - D_2$ component measure for each stimulus intensity level for Series III.

$= 3.70$, $p < .05$, Series × Seconds trends. These trends suggest that a change in wave shape developed over the three stimulus series. This result was the expected consequence of habituation to the high-intensity tones. To determine if either an error-related or intensity-related effect remained at the end of the experiment, three planned contrasts of means were performed on the cardiac wave form component measures for Series III alone. Only measure $A - D_2$ significantly differentiated stimuli, this result being extracted by Contrast 1, $F\ (1, 36) = 8.564$, $p < .01$. These results indicate that, as in the overall analysis, the two anchor stimuli elicited a larger response than the three center tones, with no significant effect remaining for intensity. Figure 4 shows the means of this measure for each tone level.

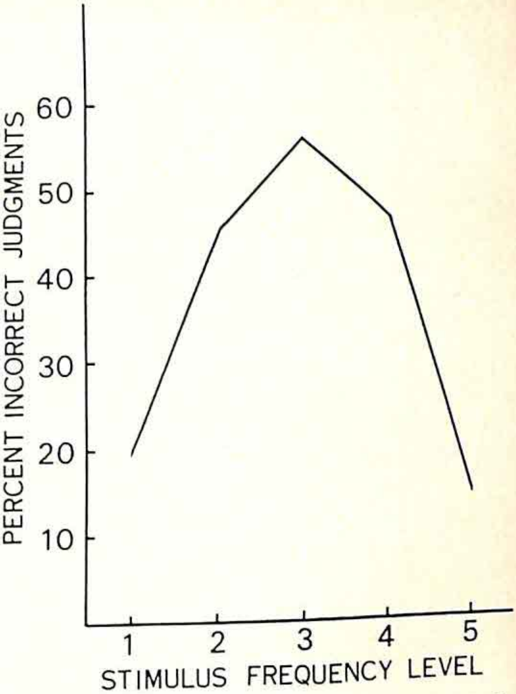


FIG. 5. Percentage of incorrect judgments for each stimulus frequency level across all series (Stimulus 5 is the highest frequency).

EXPERIMENT II

Method

Subjects. This sample again consisted of 10 male undergraduate volunteers from the introductory psychology *S* pool.

Apparatus. In addition to heart rate, slow cortical activity was continuously recorded during this experiment. Beckman silver–silver chloride electrodes were used for both measures. Electrode sites for heart rate were the same as in the previous study. EEG electrodes were positioned at the vertex and the right mastoid. A Grass Model 5PI low-level DC preamplifier, modified to introduce a 10-sec. time constant into the signal, was used to monitor the EEG responses. The output from this preamplifier was recorded on one channel of the instrumentation tape recorder.

The stimulus presentation apparatus was similar to that of Experiment I. It differed in that *E* changed both the five-position stimulus selector and the frequency dial on the oscillator to the appropriate frequency—700, 1,000, 1,100, 1,200, or 1,500 Hz., according to the semirandom schedule of stimulus presentation. To correct intensity for frequency changes, the middle tone was set for 65 db. at the site of *S*'s head. The other tones were judged to be of comparable intensity and were measured to be within 1 db. of 65 at the same location. The two speakers were placed behind *S*'s chair at floor level.

Data reduction. The HR data were reduced in the same manner as in Experiment I.

The LINC-8 was programmed for evoked potential averaging. For every *S*, curves were averaged over each series for each of the five frequency levels (10 trials/level), and also averaged across all series for each frequency level (30 trials/level). Each averaged EEG curve consisted of 300 data points distributed over a 10-sec. period, 2.03 sec. of which was prestimulus activity. The 300 points were then collapsed into 20 data values (each the average of 15 points and, thus, representing a .5-sec. sample). These values were converted to microvolts and the simplified curves were statistically analyzed on the UNIVAC 1108 computer.

Response measurement and procedure. The procedure and response measurements were similar to those of the previous experiment. The *S* was asked to rank individual presentations of five different frequencies from highest to lowest. The first tone presented was the lowest frequency.

Results

The statistical analyses performed on the judgment and heart rate data of this experiment paralleled those procedures utilized in Experiment I.

Judgmental errors. Figure 5 shows the distribution of judgmental errors across frequencies for all trials. Trend analysis indicated no basic change in the form of this curve over series, although a significant reduction in total errors was indicated: series linear component, $F\ (1, 9) = 11.29$, $p < .01$. Again, the curve shape indicates success in manipulating the discriminability of frequencies. The fewest judgmental errors were made to the anchor stimuli; the most to the middle three stimuli. This was confirmed by a large quadratic trend, $F\ (1, 9) = 50.24$, $p < .001$, accounting for almost all the variance.

Three orthogonal planned contrasts of means were performed on these judgmental errors. Contrast 1 (Stimuli 1 and 5 vs. 2, 3, and 4) was significant, $F\ (1, 36) = 80.59$, $p < .001$, substantiating the trend of the anchor stimuli being judged correctly most often. Contrast 3 (Stimuli 2 and 4 vs. 3) was also significant, $F\ (1, 36) = 8.314$, $p < .01$. However, unlike Experiment I, Stimuli 2 and 4 were judged correctly *more often* than the middle tone. As expected, there was no difference in error frequency between Stimuli 1 and 5.

Heart rate. Averages of the HR response across *S*s for the highest frequency, the lowest, and the middle tone, for Series III, are shown in Figure 6. To facilitate comparison, the curves have been brought to a common level at .5 sec. prior to stimulus onset.

A trend analysis of the data for all tones across all series indicated significant linear, $F\ (1, 9) = 11.53$, $p < .001$; quadratic, $F\ (1, 9) = 5.22$, $p < .05$; and quartic, $F\ (1, 9) = 40.95$, $p < .001$, trends across seconds. As in Experiment I, these trends reflect a triphasic cardiac response to stimulus presentation. Component measure values are presented in Table 2. The two anchor stimuli, the two stimuli judged correctly most often, appear to elicit the largest magnitude heart rate responses. Analysis of the measures by planned contrasts of means indicated that Contrast 1 was significant for the $A - D_2$ measure, $F\ (1, 36) = 4.517$, $p < .05$; the $(A - D_2) + (A - B)$ measure, $F\ (1, 36) = 5.162$, $p < .05$; and the combination measure of initial and secondary deceleration from the prestimulus base $(B - D_1) + (B - D_2)$, $F\ (1, 36) = 4.387$, $p < .05$. The first two results were present in Experiment I and support the main hypothesis: The two anchor stimuli clearly elicited larger responses than the middle tones. The last contrast result, indexing a deceleratory trend, was not present in Experiment I. It indicates that the middle three tones elicited more cardiac deceleration than the anchor stimuli.

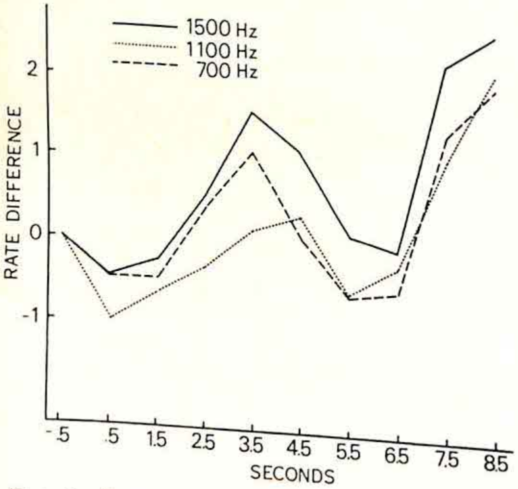


FIG. 6. Second-by-second averages of heart rate response across *S*s for the highest, lowest, and middle frequency tones for Series III—rate difference from Second − .5.

TABLE 2

MEAN HEART RATE (HR) COMPONENT MEASURES FOR EACH STIMULUS FREQUENCY LEVEL OVER ALL SERIES IN EXPERIMENT II

Stimulus frequency level (in Hz.)	HR wave form component measures (beats/minute)		
	$A - D_2$	$(B - D_1) + (B - D_2)$	$(A - D_2) + (A - B)$
700	2.70	3.00	5.50
1,000	2.10	3.40	4.30
1,100	1.90	4.80	3.25
1,200	2.25	3.20	3.75
1,500	2.75	2.60	5.00

Analysis of the raw heart rate data indicated no Series × Seconds trends which would suggest a large change in wave shape over series. However, planned contrasts of means were performed on the cardiac component measures for Series III only, in order to make direct comparisons with Experiment I. Contrast 1 was again significant for the $A - D_2$ measure, $F\ (1, 36) = 7.050$, $p < .025$; the $(A - D_2) + (A - B)$ measure, $F\ (1, 36) = 8.702$, $p < .01$; and the $(B - D_1) + (B - D_2)$ measure, $F\ (1, 36) = 5.167$, $p < .05$. In addition, Contrast 3 (Stimuli 2 and 4 vs. 3) was significant for the two combination measures $(A - D_2) + (A - B)$, F (1, 36) = 4.905, $p < .05$; and $(B - D_1) + (B - D_2)$, F (1, 36) = 6.084, $p < .025$. These results indicate that the middle tone, which was judged incorrectly most often, was associated with both the smallest magnitude heart rate wave form and the largest overall deceleration. Figure 7 plots the means of all three measures for each frequency level.

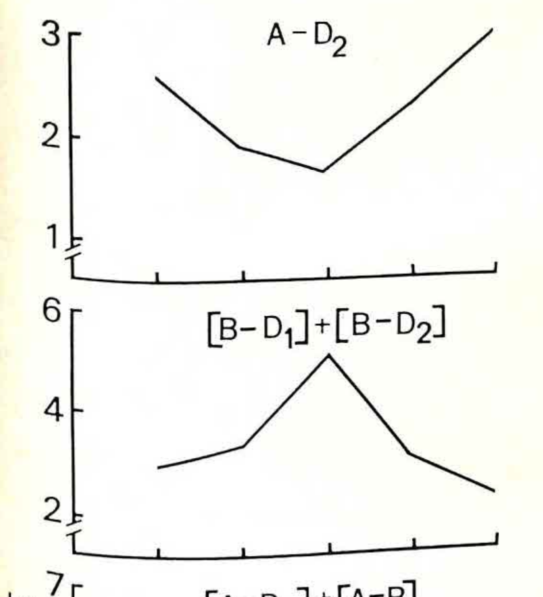


FIG. 7. Mean cardiac component measures for each stimulus frequency level for Series III.

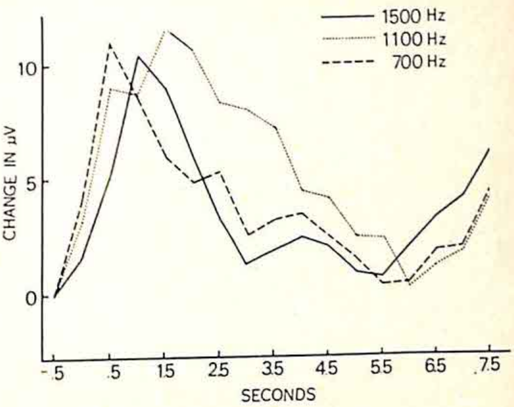


FIG. 8. Cortical slow-wave response curves across series and *Ss* for the highest, lowest, and middle frequency tones.

Cortical slow-wave activity. Figure 8 shows the averaged slow-wave curves across *Ss* and series for each frequency level. The curves have been brought to a common level at .5 sec. prior to stimulus onset. There is little evidence of the CNV (contingent negative variation) response described by Walter (1964), which suggests that *Ss* did not perceive tone offset as a temporally fixed stimulus for responding. However, it is readily apparent that there is more overall negativity associated with the middle stimulus. A trend analysis on all data indicated significant linear, quadratic and cubic trends across seconds, F_1 (1, 9) = 7.450, F_2 (1, 9) = 10.321, F_3 (1, 9) = 11.280, all ps $< .025$. Planned contrasts of means were performed on deviation scores derived from this data. Differences between the prestimulus base level and the negativity score for each of the first eight poststimulus $\frac{1}{2}$-sec. intervals were obtained. The average of these eight deviation scores was the value used in the analysis. Contrast 1 (Stimuli 1 and 5 vs. 2, 3, and 4) was significant, F (1, 36) = 4.96, $p < .05$, indicating that more negativity was elicited by the three middle tones.

Analysis of the raw data indicated a near significant quadratic Series × Seconds trend, F (2, 18) = 3.01, $p < .1$, suggesting a change in wave shape over series. Therefore, planned contrasts of means were performed on the deviation scores for Series III only. Again, only Contrast 1 was significant, F (1, 36) = 6.48, $p < .025$, indicating that more negativity was associated with the middle three, less accurately judged tones. This finding parallels the heart rate result, showing a relationship between physiological response amplitude and judgmental accuracy. However, unlike the heart rate data, the cortical slow wave did not differentiate between the middle three tones.

DISCUSSION

The results of these two experiments confirm the main hypothesis: The distribution of heart rate response amplitudes covaries with the relative discriminability of stimuli in a judgment task. When the difference between heart rate acceleratory peak and the subsequent deceleratory valley was used as an index (Lang & Hnatiow, 1962), larger magnitude responses were associated with the two consistently identified anchor stimuli. The relatively small sample of *S*s required to demonstrate this phenomenon attests to its strength and statistical reliability.

As previously noted (Bull & Lang, 1972), tone intensity also determined response amplitude in the first experiment. However, this factor appeared to habituate over series, and by Series III only the discriminability of tones still modulated the heart rate wave form.

In the second experiment, when stimuli were varied in frequency rather than intensity, an even closer parallel between judgment accuracy and heart rate magnitude was observed. Thus, the tone occasioning the most errors (the center frequency) also yielded a significantly smaller peak to valley response ($A - D_2$) than did the slightly more accurately judged adjacent tones. This effect was noted during Series III, but not in the overall analysis. It would appear that the finer grain of the stimulus judgment–heart rate relationship is evident only after *S*s have considerable familiarity with stimulus materials.

The psychophysiological significance of these results is not completely clear. In the Bull and Lang (1972) study, a simple reaction time experiment was superimposed on the judgment task. Rapidity of response covaried with judgment errors in the same fashion as in studies of response certainty (Garner, 1962). It was hypothesized that the stimulus-evoked heart rate changes were secondary to reaction speed, part of an overt response preparation process. This seems less likely in the present experiment, in which there was no RT task and no encouragement to make judgments rapidly.

The present data could mean that over trials and series, well discriminated stimuli take on signal properties. That is, they become conditioned stimuli for specific verbal–cognitive responses (e.g., highest and lowest). Thus, as in conditioning studies (Dawson, 1953; Geer, 1964), we observe an augmented cardiac wave form to these signal tones. From this perspective, the smaller wave forms found for the middle tones could represent a simple failure to respond. These tones never prompt reliable judgments and thus less clear conditioning of specific cardiac changes.

Alternatively, it could be argued that, while the middle tones do not show as large an overall wave form, they do show a significant change in cardiac rate. Furthermore, the largest components of these responses are different from those obtained for better discriminated stimuli. In the second experiment, deceleration from base line was significantly greater for the least accurately judged tones. This result is similar to the findings of Wilson and Duerfeldt (1967) in a study of cardiac responsiveness to conflict. The *S*s in their experiment were required to press one of two buttons, indicating whether a stimulus light matched a training stimulus that signaled electric shock, or was more like a no-shock light. More difficult discriminations were associated with greater secondary deceleration. Graham and Clifton (1966) and Lacey and Lacey (1970) hold that cardiac deceleration is a correlate of stimulus receptivity and orienting. In both our experiment and that of Wilson and Duerfeldt (1967), the difficult stimuli may have prompted more concentrated attention from *S*s and thus a slower heart rate, which the more readily identified stimuli did not demand. Unfortunately, the latter investigators employed a less refined heart rate analysis than in the present study. Perhaps for this reason the augmented acceleratory component of the wave form (associated with discrimination accuracy in our study) was not apparent in their data. Thus, further investigation will be needed to determine if the two experiments are truly in harmony, or whether differences in the kinds of judgment tasks used are prompting interesting differences in the evoked heart rate response.

The results for the slow EEG potential are similar to those obtained for the deceleratory heart rate components. It is the middle stimuli in Experiment II which occasioned significantly greater cortical negativity, and Walter (1964) and others (Tecce, 1972) have argued that such negativity is associated with attention. Furthermore, the absence of a contingent rising negativity prior to responding (CNV) argues that this increment is not anticipatory of an overt response, but is more likely related to the perception of the tone stimulus itself. A similar early negative change has recently been reported by Weerts and Lang (in press). Their data showed that this phenomena habituates over trials on an RT task requiring no stimulus judgments.[4]

In summary, the present data suggest close parallels between specific physiological events and cognitive responses to stimuli. The slow cortical change, and particularly the evoked heart rate wave form, appear to be good tools for studying the discriminability of stimuli, and perhaps also the certainty of *S*s' judgmental responses.

---

[4] Vertical eye movement was also recorded during the present experiment, and the data were reduced "on-line" through a Technical Measurements Corporation computer of average transients. Apparatus limitations prevented the systematic analysis of this measure across all experimental variables. However, the average poststimulus wave form of the eye movement response did parallel the overall average pattern of poststimulus cortical negativity (CNV). Further work will be needed to establish that the relationship shown here, between cortical events and judgmental responses, is not mediated by differential eye movements. Results from our laboratory relevant to this methodological issue were reported by Weerts and Lang (in press).

REFERENCES

BULL, K., & LANG, P. J. Intensity judgments and physiological response amplitude. *Psychophysiology*, 1972, **9**, 428–436.

CONNOR, W. H., & LANG, P. J. Cortical slow-wave and cardiac rate responses in stimulus orientation and reaction time conditions. *Journal of Experimental Psychology*, 1969, **82**, 310–320.

DAWSON, H. E. Concurrent conditioning of autonomic processes in humans. Unpublished doctoral dissertation, Indiana University, 1953.

GARNER, W. R. *Uncertainty and structure as psychological concepts.* New York: Wiley, 1962.

GEER, J. H. Measurement of the conditioned cardiac response. *Journal of Comparative and Physiological Psychology*, 1964, **57**, 426–433.

GRAHAM, F. K., & CLIFTON, R. K. Heart rate change as a component of the orienting response. *Psychological Bulletin*, 1966, **65**, 305–320.

LACEY, J. I., & LACEY, B. C. Some autonomic-central nervous system interrelationships. In P. Black (Ed.), *Physiological correlates of emotion.* New York: Academic Press, 1970.

LANG, P. J., & HNATIOW, M. Stimulus repetition and the heart rate response. *Journal of Comparative and Physiological Psychology*, 1962, **55**, 781–785.

TECCE, J. J. Contingent negative variation (CNV) and psychological processes in man. *Psychological Bulletin*, 1972, **77**, 73–108.

WALTER, W. G. Slow potential waves in the human brain associated with expectancy, attention, and decision. *Archiv für Psychiatrie und Zeitschrift für die gesante Neurologie*, 1964, **206**, 309–322.

WEERTS, T., & LANG, P. J. The effects of eye fixation and stimulus and response location on the contingent negative variation (CNV). *Journal of Biological Psychology*, in press.

WILSON, R. S., & DUERFELDT, P. H. Cardiac responsiveness and differential conditioning. *Journal of Comparative and Physiological Psychology*, 1967, **63**, 87–94.

(Received June 22, 1972)

Journal of Experimental Psychology
1973, Vol. 98, No. 1, 184–195

EYE MOVEMENTS DURING VISUAL SEARCH AND MEMORY SEARCH[1]

JOHN D. GOULD[2]

IBM Research Center, Yorktown Heights, New York

Nine *S*s first fixated a set of 1, 2, or 3 standard (memory) alphabetic characters and then scanned 12 comparison characters for one that matched a standard. Effects of surrounding standard characters and comparison characters with visual noise were also evaluated. Visual noise exerted independent effects upon standard set characters and upon comparison set characters. In both cases, encoding times, as measured by fixation durations, increased with the addition of visual noise, but comparison times were unaffected. Fixation durations increased on comparison characters when *S*s had to search for more than one standard, as Sternberg's 1967 theory would predict. Memory search rate was about the same on each successively fixated comparison character. Time to encode each comparison character was proportional to its vertical distance from the previously fixated one.

In visual search, studied long ago by Jastrow (1892) but more influentially by Neisser (e.g., Neisser, Novick, and Lazar, 1963), *S*s make a series of eye fixations during which they look for one or more target items on a display. These targets are memorized and have been variously referred to as the target set, the standard set, or the memory set. The *S* visually searches until he finds a target or until he decides that no target is present. Thus, visual search is thought to be self-terminating.

In memory search, which probably occurs during each eye fixation, *S*s must compare the fixated item(s) with each of the standard or memory set items to determine whether the fixated item matches one of them. Sternberg (1967, 1969) has inferred that memory search consists of two successive stages: an *encoding* stage, in which the fixated item is coded into a mental representation, and a *comparison* stage, in which the newly encoded item is compared successively with each of the previously encoded memory set items. Memory search is thought to be serial and exhaustive, i.e., each fixated item is always exhaustively compared with every item in memory, even in the cases in which the fixated item matches a memory set item.

Although eye movements have been previously recorded in visual search tasks (e.g., Gould, 1969, for a review; Kaplan, Carvellas, & Metlay, 1966; Williams, 1967), the present study is the first to measure the duration of fixations on each item separately, hence providing direct evidence about visual search and memory search rates on an item by item basis. The *S*s were shown a display of 1–3 standard characters (see Figure 1). After fixating these, they then searched 12 comparison characters for a target (i.e., a character that matched one of the standards). It was assumed that records of fixation durations, unlike overall reaction times, allow direct estimates of the processing time of each item in the standard set and each item in the comparison set. Similarly, sequences of fixations and refixations provide a basis for inferences about *S*'s attentional sequence.

Among the specific questions investigated were the following. Will the duration of fixations on individual comparison set items increase, as would be predicted from the findings of Nickerson (1966) and Sternberg (1967), when *S*s look for more than one standard? Do *S*s continually fixate each successive comparison set item, as has been assumed by others, or do they look back? Does encoding time remain constant throughout a series of eye fixations, or does it systematically speed up or slow down? Does memory comparison rate remain constant throughout a series of eye fixations, or does it undergo some systematic change, e.g., gradually change from serial to parallel?

---

[1] This research was supported by and conducted at the IBM Research Center, Yorktown Heights, New York, in 1969. The author thanks Louella Cook for help in data collection and data analyses. The preparation of the manuscript and some of the data analyses were accomplished with the support of a National Institute of Mental Health Grant MH-07722 while the author was a Visiting Associate Professor in the Department of Psychology at Carnegie-Mellon University. The author thanks William Chase and Stephen Ellis for reading an earlier version of this manuscript and providing valuable suggestions.

[2] Requests for reprints should be sent to John D. Gould, IBM Research Center, P. O. Box 218, Yorktown Heights, New York 10598.

If the comparison set characters are noisy, how will this affect eye fixations? Comparison set characters were made noisy by surrounding them with irrelevant characters, analogous to Mackworth (1965), who concluded that identification errors increased dramatically with such noise. Sternberg (1967) and Chase (1969) found that noise led to an increase in time required to *encode* a noisy (comparison) item but had little or no effect on the time to *compare* it with standard (memory) items. This result leads to the prediction of longer fixation durations on noisy comparison characters than on non-noisy comparison characters, with these longer fixation durations reflecting longer encoding times and no modification of the memory search rate.

If the standard characters are made noisy, how will this affect eye movements? No previous search study has assessed this variable. One possibility is that noise around the standard set items, unlike noise around the comparison set items, affects memory search rate. This would suggest that, whereas comparison set items are mentally encoded in a "pure" list or form prior to the comparison process, the standard set items are not. Another possibility is that noise in standard set items, analogous to noise in comparison set items, affects only their encoding. This would suggest that two pure representations are compared in memory search.

What is the rate of processing target and nontarget items in the comparison set? Neisser (Neisser et al., 1963) has studied the processing rate of nontargets only, whereas Sternberg and Scarborough (1969) in their "simulated search" experiment were able to assess the processing rate of both targets and nontargets but only on positive trials.

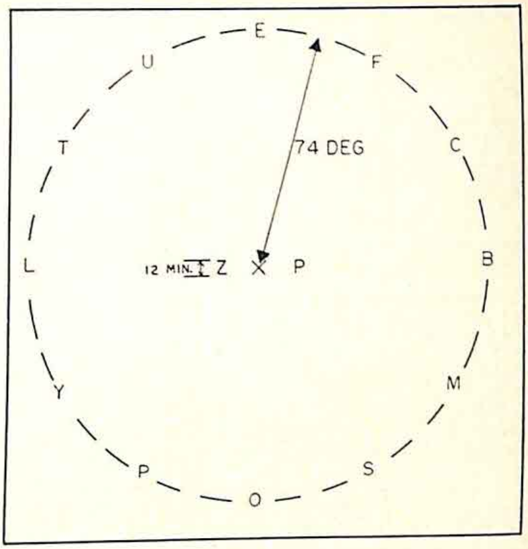


FIG. 1. A schematic stimulus.

METHOD

In the center of a stimulus, shown schematically in Figure 1, were 1, 2, or 3 alphabetic characters, called the standard set. Equally spaced on an imaginary circle, centered on the middle standard character and of radius 7.4° angular subtense, were 12 comparison set alphabetic characters. The characters were computer generated and photographed as 35-mm. negative slides which resulted in white-on-black stimuli. When back projected on a screen in front of *S*, each character had an angular subtense of about 12′ at *S*'s eye. The luminance of the background was about 1 mL. and the luminance of the characters (which of course varied from character to character and in the usual way across the screen) was about 10 mL. At any place on the screen, characters had roughly 10 times more luminance than adjacent background. The standard set and comparison set characters were changed on every stimulus. The nontarget comparison set characters were randomly generated.

Each of nine paid college *S*s was instructed to first fixate the standard set character(s) and then to fixate the comparison characters, starting at the 12:00 position and moving clockwise. Each *S* was instructed to find, as fast as he could, subject to being accurate, the one comparison set character that matched a standard set character, i.e., the target. There was always one target on a stimulus. Refixations were allowed. Upon finding the target, *S* pushed a button that caused a shutter to occlude the stimulus. He then indicated the target position by holding up the appropriate number of fingers, and was given correct-answer feedback.

Standard set character(s) were either not noisy, embedded in the center of a 5 × 5 matrix of irrelevant alphabetic characters, or embedded in the center of a 9 × 9 matrix of irrelevant alphabetic characters. The vertical subtense of each noise character was also about 12′. The vertical subtense of the 5 × 5 array was about 1°24′ and the vertical subtense of the 9 × 9 array was about 2°36′. Noise characters varied from one stimulus to another. Comparison set items were also either not noisy, embedded in the center of a 5 × 5 matrix of irrelevant characters, or embedded in the center of a 9 × 9 matrix of irrelevant characters. The target item could appear at any of the 12 clock positions in the comparison set.

Thus, there were 324 conditions in this experiment (3 standard set sizes, 3 standard set noise levels, 3 comparison set noise levels, 12 target positions), and each *S* went through all of them over a period of 5 days. Each of the 27 conditions of main interest (standard set size, standard set noise, and comparison set noise) was seen an equal number of times on each day. The variable of least interest, target position, was spread across days in a manner that was different for each *S*. On the first day, each *S* received four blocks, at various target positions, of these 27 conditions. On each of the next 4 days, each *S* received two more blocks of these 27 conditions at various different target positions. Each *S*'s subjective uncertainty as to target position was always 1/12 on every stimulus. Order of presentation within a block was random.

Overall reaction time was recorded on all days. Eye movements were recorded on the last 4 days but, because of the amount of labor involved, the results were analyzed on the last 2 days only.

The design was so arranged that, on the last 2 days, $\frac{1}{4}$ of the 108 stimuli that each *S* viewed were from the first three target positions, $\frac{1}{4}$ were from the second three target positions, etc. The results of a group of three *S*s on the last 2 days provided one replication of the 324 different conditions. The three groups of three *S*s provided three such replications.

Eye movements were recorded with a system that produced videotaped records of the stimulus field and the positions of *S*'s fixation within it. This was accomplished with a modified closed-circuit television, corneal-reflectance, eye-marker system akin to that developed by Mackworth and Mackworth (1958) and described previously (Gould & Schaffer, 1965). A television camera, securely positioned above *S*'s head, transmitted to a closed-circuit television monitor a picture of the projected stimulus. Superimposed upon this picture on the monitor was a small light spot or eye marker (about 1°) whose location corresponded to the position of *S*'s fixation on the projected stimulus. To accomplish this, at the vidicon tube of the camera the virtual image, formed by the cornea, of a small light source was optically superimposed on the image of the stimulus field. With *S*'s head fixed in place, the position of this image was proportional to the angle of deviation of the fovea from the primary position.

Eye movements were recorded to an accuracy sufficient to determine which character position *S* was looking at most directly with each fixation. When all precautions are taken, the accuracy of recording is roughly 1°. Preexperimental work showed that such refinements (e.g., using a calibration matrix to correct the data of each *S*) were not required. When *S* fixated a noisy character, it was theoretically possible, for example, that he could sometimes make two separate, relatively brief eye fixations on different parts of it and that this would be detected as one relatively long eye fixation. Since recording accuracy refers to the ability to be sure where a person is looking, and not to the smallest movement possible, such movements would usually be detectable. Compared to other anomolies in the data, they seemed to occur only rarely and were ignored. The main anomoly was that, from time to time, the reflected light spot would disappear from the record and then reappear in the same character position. It was assumed that *S* had closed his eye. In calculating the time for this fixation, this blink time plus the times the light spot was in the character position were included. When a stimulus was first projected *S* usually required some time to move his eye from the fixation point to the standard set. This time was included in overall search time results, of course, but was not included in the times of fixating individual items.

Prior to the start of each trial, *S* fixated a point a few degrees northeast of where the standard set would be projected. This allowed *E* to do any necessary calibration prior to the onset of the next stimulus. The *S* was then told "ready" and a new stimulus was projected. Several seconds usually elapsed between *S*'s response to one stimulus and the presentation of the next one.

The duration of each fixation (recorded at 60 frames/sec on videotape, an accuracy of ± 8 msec.) was read from the taped records and, together with a code describing the particular pattern fixated, punched into an IBM card. This process was repeated for each additional fixation on that trial, resulting in an ordinal record suitable for computer processing of the duration, location, and number of fixations on each stimulus.

Except where specified otherwise, the results for a particular eye-movement parameter (e.g., mean number of fixations on standard characters, mean duration per fixation on comparison characters) were obtained by calculating the average value of that parameter for each trial and then doing an analysis of variance on these 972 resulting scores.

RESULTS

Errors. There were few errors (less than 1%). An error occurred when *S* reported the position of the target incorrectly.

Time scores. Only a general overview of overall search times will be given; the main data and detailed analyses will concentrate

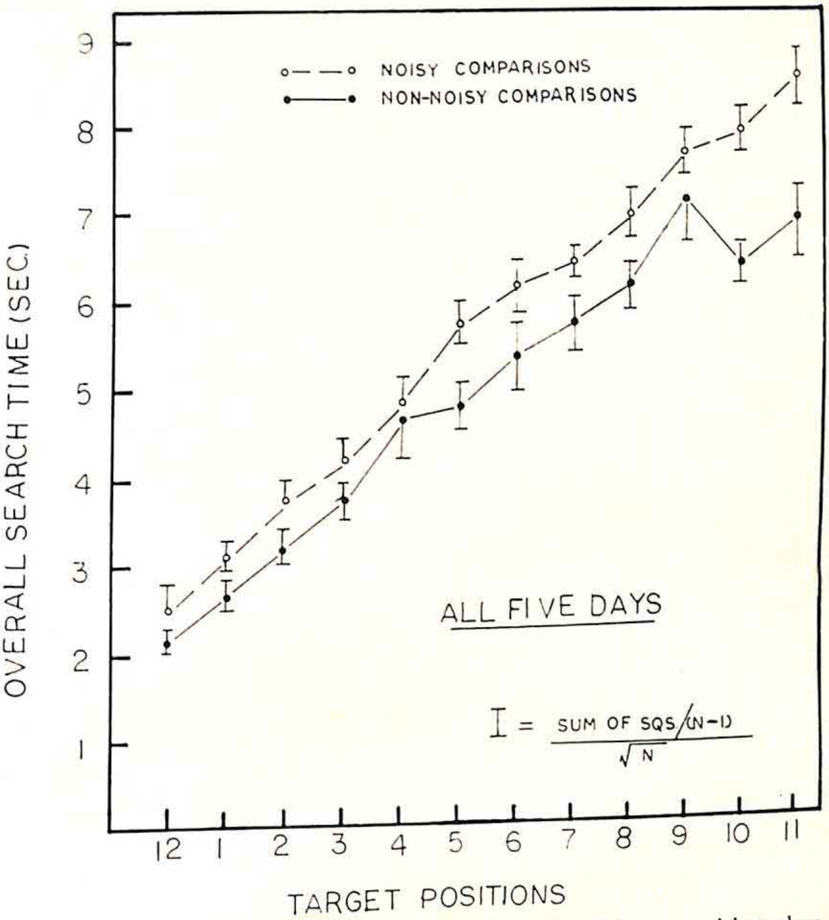


FIG. 2. Overall search times on 5 days as a function of target position when the comparison characters were not noisy (solid curve) and when the comparison characters were noisy (broken curve).

on eye movements. As can be seen in Figure 2, overall visual search times increased monotonically with target position around the clock, F (11, 88) = 135.6, $p <$ .001, similar to Neisser's (Neisser et al., 1963) finding of a linear increase for lists of characters.

Visual search times became longer when the comparison characters were noisy; overall means for the three comparison set noise levels were 4.93, 5.62, 5.82 sec., respectively; F (2, 16) = 213.38, $p <$ .001. Figure 2, which combines the data for the two noise levels, shows that this difference became increasingly greater as *S*s had to search more and more comparison item positions, i.e., a significant Target Position × Comparison Set Noise interaction, F (22, 176) = 2.36, $p <$ .01. Results of eye movement analyses explain this interaction below. The widening spread in the curves, which occurs as *S*s look at more and more comparison characters, was due to the fact that fixation durations were longer on noisy comparison characters than on non-noisy comparison characters. As more characters of each type were fixated, the spread became greater.

Overall search times increased when the standard set characters were noisy also; overall means for the three standard set noise levels were 5.15, 5.50, and 5.71 sec., respectively; F (2, 16) = 18.47, $p <$ .001. However, as the eye-movement data will show, visual search rate was unaffected by noise around the standard set characters.

Overall search times increased significantly with the number of standards; overall means for the three standard set sizes were 4.31, 5.42, and 6.64 sec., respectively;

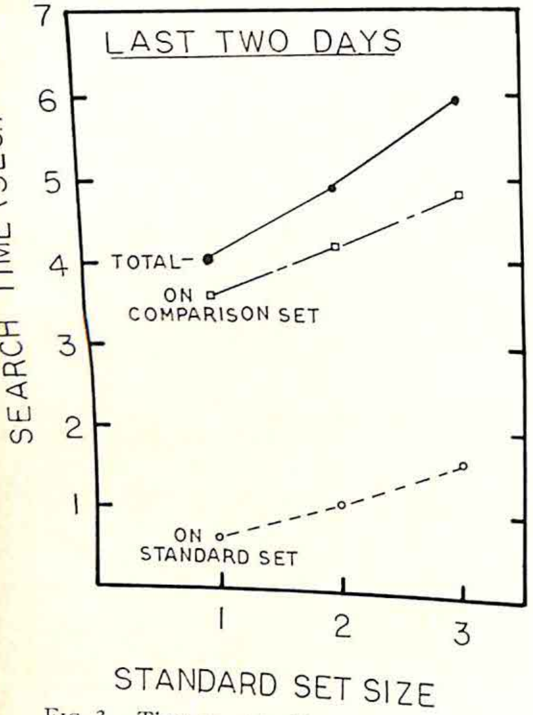


FIG. 3. Time spent looking at standard set and comparison set characters.

F (2, 16) = 79.15, $p < .001$. As shown in Figure 3 the overall search times on the last 2 days were divided, on the basis of eye-movement records, into the amount of time *S*s spent looking at the standard set and the amount of time *S*s spent looking at the comparison set. The time spent looking at standard items (lowest curve) increased by about 550 msec. for each additional standard set item. This suggests that a constant amount of time was required for encoding and remembering each standard character. The time spent looking at comparison set characters increased by about 600–650 msec. for each additional standard set item. This suggests a sequential memory search process of constant rate. On the average, there were 5.5 comparison characters that preceded the target character. If *S*s fixated each of these and the target character once and only once, this suggests a memory search rate of about 100 msec. per item. As will be seen however, fixation durations on comparison set items increased by only about 50 msec. per standard set item. Thus, as described below, the order of visual search also accounted for part of this apparent 100-msec. memory search rate.

The overall search times are about 10% greater than the sum of the fixation durations on standard set and comparison set characters (solid curve, Figure 3). This discrepancy is due to the fact that overall search times include the fixation durations on the fixation point prior to *S* moving his eye initially, the travel times of the eye, and the infrequent occasions when it was impossible to tell where *S* was looking.

Fixations on standard (memory) characters. Each *S* usually fixated each standard with a single, separate eye fixation. Noise around either the standards and/or around the comparison set characters did not cause refixations of standards. The number of refixations of the standard set was affected by the number of standards present, however, F (2, 16) = 15.06, $p <$ .01. As shown in Figure 4, the probability of refixating the standard set during a trial increased 3%–38% as the standard (memory) set size increased 1–3. There were 8 refixations of the standard set on the 324 stimuli that contained one standard, whereas there were approximately 30, 62, and

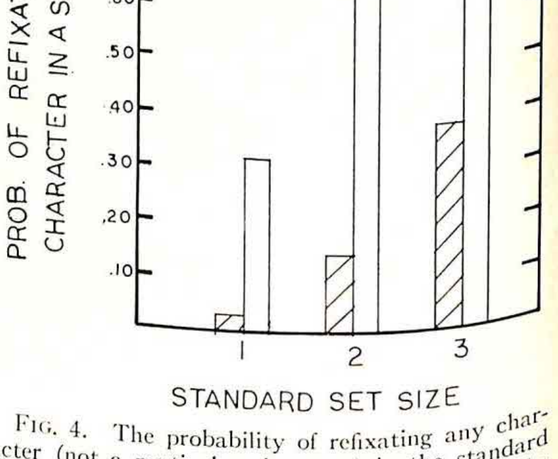


FIG. 4. The probability of refixating any character (not a particular character) in the standard set and in the comparison set as a function of the standard (memory) set size.

29 refixations on the left, center, and right standards, respectively, on the 324 stimuli that contained three standards. Thus, the explanation for at least part of longer visual search times with increasing memory load is based upon this change in visual scanning behavior. This result is in contrast to that of Kaplan, Carvellas, and Metlay (1966) who found, in a different kind of task, a constant number of glances between a subset and a containing set when the subset had 1–4 targets.

Overall, mean fixation duration on standards was 514 msec. Whereas the total time spent looking at the standard set increased with standard set size, as reported above, the duration of eye fixations on *each* standard decreased as the number of standards increased (see Figure 5), F (2, 16) = 45.31, $p < .001$. When there was one standard, fixation durations averaged 600 msec. When there were three standards, fixation durations averaged 472, 431, and 460 msec. on the left, center, and right standards, respectively. These values include the length of time needed to process each standard set character and to prepare to move the eye to the comparison set characters. They probably reflect very little processing of the comparison set characters in peripheral vision because (*a*) the acuity required for this peripheral shape processing to be useful is reported to be insufficient (Feinberg, 1949; Mandlebaum & Sloan, 1947; Williams, 1967), and (*b*) there is no selectivity in the subsequent fixation sequence on comparison set characters to indicate such processing had occurred. Such selectivity has been found for gross shape processing (Gould & Dill, 1969).

The durations of fixations on standard set characters were much longer when the standards were surrounded with noise than when they were noiseless; two noise conditions are combined in Figure 5, F (2, 16) = 88.71, $p < .001$. Taken together with the fact that noise around the standard set had no effect upon fixation durations on comparison set characters (discussed below), this suggests that noisy standard set items, after encoding, are available in some "pure representation" for comparison.

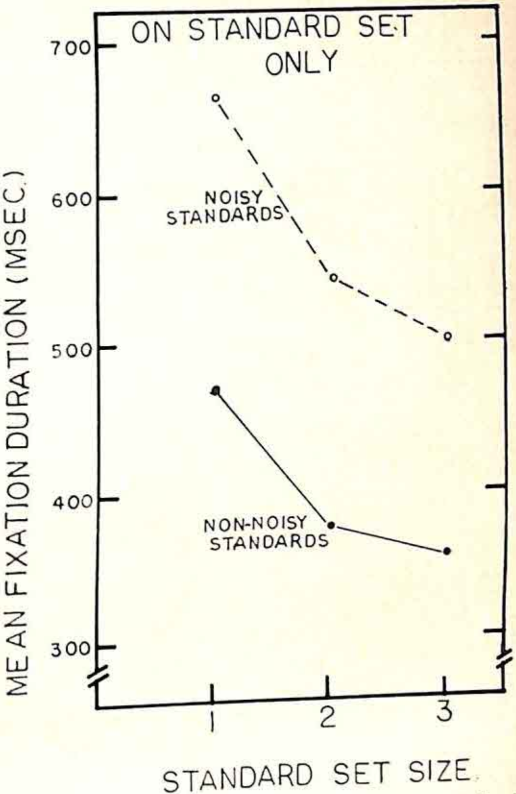


FIG. 5. Mean fixation durations on each standard set character as a function of standard set size when the standard set characters were not noisy (solid curve) and when they were noisy (broken curve).

The durations of fixations on standards were not affected by noise around the comparison set characters. This suggests that the standard set and the comparison set were processed independently.

Fixations on comparison items. The *S*s first fixated the comparison character in the 12:00 position and, in general, then fixated each other comparison set character once and only once, stopping when they reached the target. Variation in the number of refixations was primarily due to variation in standard set size, F (2, 16) = 6.24, $p < .01$. When there was one standard set item there was about a 1 in 3 chance that *S*s would look back at one comparison set item (Figure 4). That is, there were 103 refixations on the 324 stimuli that contained one standard. As shown in Figure 4, when the standard set increased to three, there was a virtual certainty that

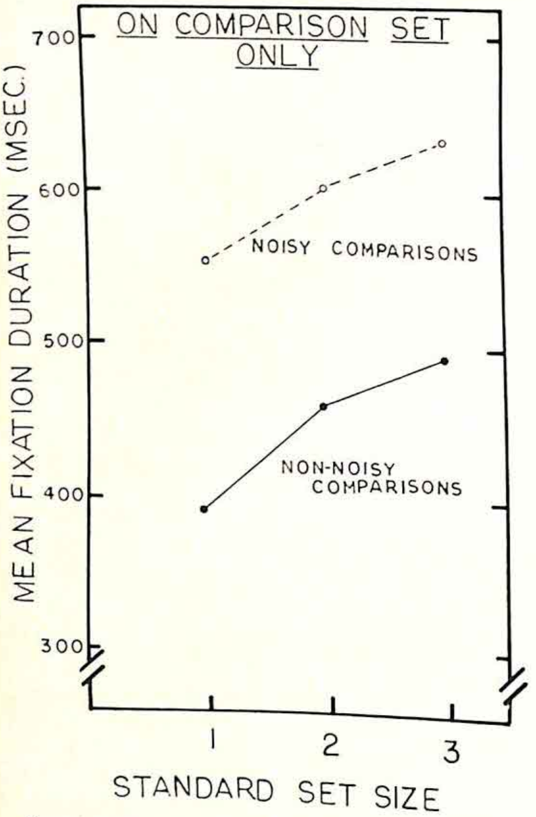


FIG. 6. Mean fixation durations on noisy (broken curve) and non-noisy (solid curve) comparison set items as a function on standard set size.

one comparison character would be refixated on each stimulus. There were 331 refixations on the comparison characters on the 324 stimuli that contained three standards. This indicates that refixations of comparison set characters were a second reason for increased visual search time with an increase in memory load.

Figure 4 also shows that comparison set characters were more likely to be refixated than standard set characters. Refixations on the standard set were probably required because *S* either did not attend carefully enough to the identity of these characters initially or else he subsequently forgot their identity. The greater number of refixations on the comparison set was partially due to (*a*) the presence of more comparison set items than standard set items on the average stimulus, and (*b*) occasional looking back again after completing a scan and not finding a target. In addition, some refixations on comparison set items *may* have been due to (*a*) the eye overshooting, and (*b*) *S*s' lack of attention when fixating an item originally because of rapid visual scanning.

The *S*s sometimes scanned past the target character and then returned to it. There were 57, 87, and 119 of these "extra" fixations on comparison set characters for standard set sizes of 1, 2, and 3, respectively.

Of all fixations on comparison set characters, 1,134 were on the 972 targets shown and 6,023 were on the nontarget comparison characters. Based upon all the stimuli, 5,346 nontargets preceded the target characters.

Figure 6 plots the mean duration of eye fixations on the comparison set as a function of the number of characters in the standard set. The two noise conditions are combined. The mean durations of eye fixations on comparison characters were longer when they were noisy than when they were not noisy, F (2, 16) = 136.31, $p < .001$. At each standard set size, fixation characters were about 150 msec. longer on noisy comparisons than on non-noisy comparison items. That is, the main effect of noise was to raise the intercept of the curve about 150 msec. but not to affect the slope of the curve. This result, based upon eye-movement reaction time, is consistent with Sternberg's (1967) and with Chase's (1969) conclusions, based upon finger reaction times, that noise affects stimulus encoding but does not affect memory comparison. Taken together with fixation data on standard set items, it also suggests that memory comparison consists of the comparison of two pure representations or feature lists, both of which have been stripped of their noise during encoding.

As shown in Figure 6, fixation durations on comparison set items increased by about 30–50 msec. for each additional item in the standard set, F (2, 16) = 34.44, $p < .001$. This slope is interpreted by Sternberg (1967) as the memory search rate. The increase in time is consistent with Sternberg's serial memory scan model as well as with some parallel models; in this experiment it is a third reason for longer

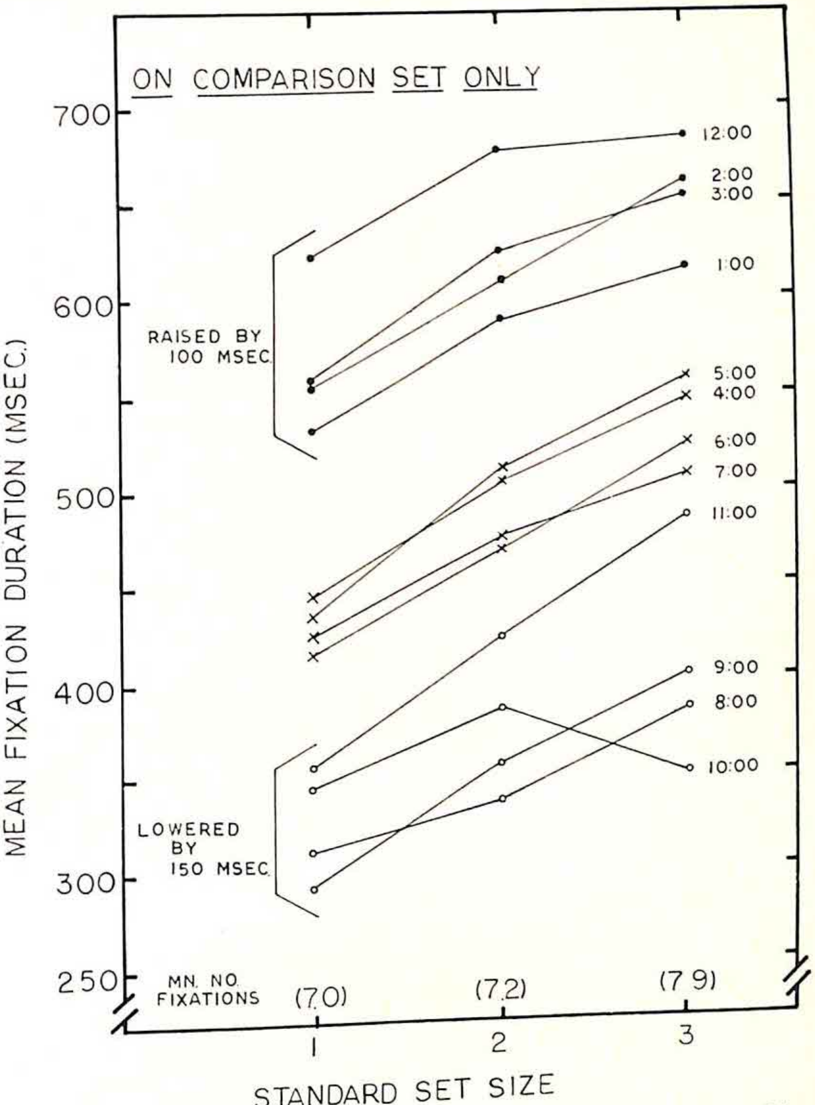


Fig. 7. Mean fixation durations on comparison set items at each clock position.

visual search times with an increase in standard set size.

In Figure 7 the mean fixation durations on each successive position of the comparison set item are plotted as a function of standard set size. The top four curves, which are the fixation durations on the comparison set characters in the 12:00, 1:00, 2:00, and 3:00 positions, have been arbitrarily raised 100 msec. The bottom four curves, which are the fixation durations on comparison set characters in the 8:00, 9:00, 10:00, and 11:00 positions, have similarly been lowered 150 msec. The curves appear roughly parallel. The significance of this is that it suggests memory search rate is nearly the same on each successive comparison set item. It does not, for example, change from serial to parallel as *S* fixates additional comparison set characters, as would be suggested if the curves at the later clock positions had slopes near zero.

Figure 7 shows another important finding. The (projected) intercepts of the 12 curves are roughly the same. This indicates that as *S* fixates successive comparison set items there is no systematic decrease in eye fixations, as Llewelyn-Thomas and Lansdown (1963) reported for the scan-

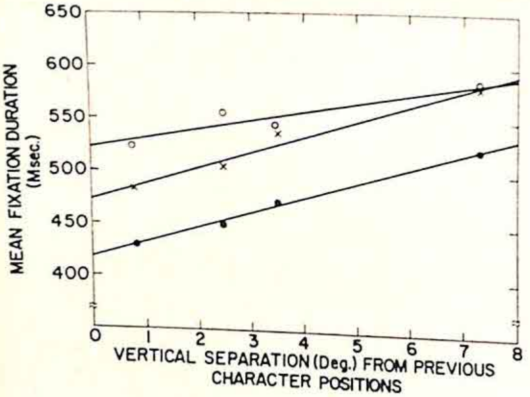


FIG. 8. The relationship between the vertical distance to a character position and how long that position was fixated for each standard set size.

ning of X rays. Likewise, there is no systematic increase either. Based upon Sternberg's (1967) and others' models of item recognition, the intercept of each curve includes the time required to encode a fixated item. Thus, encoding time did not systematically increase or decrease throughout the series of eye fixations.

There were some differences, however, among the projected intercepts of Figure 7. These differences were almost entirely accounted for by the fact that the vertical separations of characters equally spaced around a circle are different. There were four categories of vertical separation (midpoint of one character from midpoint of character in previous position): Positions 1, 6, 7, and 11 (.8°); Positions 2, 5, and 8 (2.5°); Positions 3, 4, 9, and 10 (3.5°); and Position 12 (7.4° from midpoint of center standard). Figure 8 plots the mean fixation duration on comparison set items as a function of each of these four separations. The lowest curve, based upon a standard set size of one, shows a linear relationship between mean fixation duration on characters and the vertical separation of their position from the previous character position (d). The equation $y = 14.5\ d + 417$ msec. accounts for 99% of the variance among the fitted points. This suggests that while S fixated one character he encoded information about the next character in proportion to its vertical separation from the one he was fixating. This information could be either positional, which might reduce any ocular adjustments required to bring the next character into clear view once it was fixated (cf. Bartz, 1967), or it might be gross geometric features, which might reduce the required amount of subsequent feature processing once the next character was fixated.

This linear relationship was not as great when *S*s were matching comparison set items against two or three standards. For two standards (Figure 8), the best-fitting linear curve accounted for 97% of the variance among the four fitted points, whereas the best-fitting quadratic function accounted for slightly more variance (98%). For three standards, the best-fitting linear function accounted for 87% of the variance among the four fitted points, and the best-fitting quadratic function also accounted for 87% of the variance.

This suggests that with one standard, S can restrict his attention to the comparison set characters, whereas with two or three standards the additional rehearsal and other processing requirements wash out this relationship.

The slopes in the curves of Figures 6 and 7 appear greater when standards increased from one to two than when standards increased from two to three. Table 1, which shows the mean numbers and durations of fixations at each pattern position for each standard set size, indicates that this was due to a trade-off between fixation durations and numbers of eye fixations.[3] When there were one or two standards, fixation durations increased by about 50 msec. and number of fixations remained about the same. When there were three standards, however, fixation durations increased by about only 30 msec., but now there were about 10% more eye fixations. The important point is that if total time is taken as the criterion, then memory search rate

---

[3] The weighted mean fixation durations of Table 1 are about 50 msec. less than those obtained by calculating the mean fixation duration on comparison characters on each trial and then averaging these means, as presented above. This is so because, as can be seen in Table 1, some of the early pattern positions, especially the 12:00 position, that are fixated on most trials are longer than some of the less frequently fixated positions.

TABLE 1

MEAN FIXATION DURATIONS (IN MSEC.) AND NUMBER OF FIXATIONS (IN PARENTHESES) ON EACH CHARACTER POSITION FOR STANDARD SET SIZES OF 1, 2, AND 3

Character position	Number of positions to be searched	Standard set size		
		1	2	3
Standard set				
Left	324	—	503(348)	470(354)
Center	324	608(326)	474(348)	426(386)
Right	324	—	—	457(353)
Comparison set				
12:00	324	523(336)	579(342)	583(365)
1:00	297	433(309)	489(315)	514(338)
2:00	270	453(285)	509(293)	561(310)
3:00	243	456(261)	524(258)	553(283)
4:00	216	446(231)	505(229)	548(255)
5:00	189	436(198)	512(205)	559(228)
6:00	162	415(177)	469(187)	524(200)
7:00	135	425(148)	475(154)	509(164)
8:00	108	463(126)	487(123)	535(149)
9:00	81	442(92)	508(108)	551(116)
10:00	54	494(64)	536(78)	500(92)
11:00	27	505(36)	573(46)	637(54)
Weighted mean duration on comparison set		456	513	546

Note. Mean number of fixations on comparison set were 7.0, 7.2, and 7.9, for standard set sizes of 1, 2, and 3, respectively.

appears constant (cf. middle curve, Figure 3). Table 1 shows, however, that a revised order of visual scanning, including more refixations and more extra fixations (described above), occurs with an increase in standard set size. No simple serial memory-scan model portrays this.

Whether or not the standard characters were noisy had no effect upon fixation durations on comparison set characters. Mean fixation durations on comparison characters were 551 msec. when standards were not noisy and 548 msec. when standards were noisy. As mentioned above, this suggests that visual noise affects the encoding times of standards but not the comparison times.

Some informal analyses indicated that context effects (of various different noise characters on various different standard and comparison characters) exerted a major effect upon eye movements, just as Neisser et al. (1963) found for visual search times.

The duration of fixation on nontarget comparison characters (448 msec.) was considerably shorter than on target comparison items (693 msec.), a result similar to that found previously for histoform patterns of various complexities (Gould & Dill, 1969). Figure 9 shows that fixation durations on both targets and nontargets increased with increased standard set size. The curves are nearly parallel, which is consistent with Sternberg's (1967) interpretation of exhaustive memory scanning.

DISCUSSION

The finding that visual noise around comparison items affects their encoding times but not their comparison times replicates previous work (Chase, 1969; Sternberg, 1967), but the findings (*a*) that the same result occurs for standard set items and (*b*) that noise exerts independent effects upon standard (memory) items and upon comparison items extend previous understanding. Memory search rate is not affected by visual noise around either standards or around comparison characters. Of course, it would have been possible for noise around the standard characters, for example, to have affected memory search rate even though noise around the comparison set did not. It appears that on structured displays such as used here, *S*s eventually extract and retain (albeit perhaps briefly) the same infor-

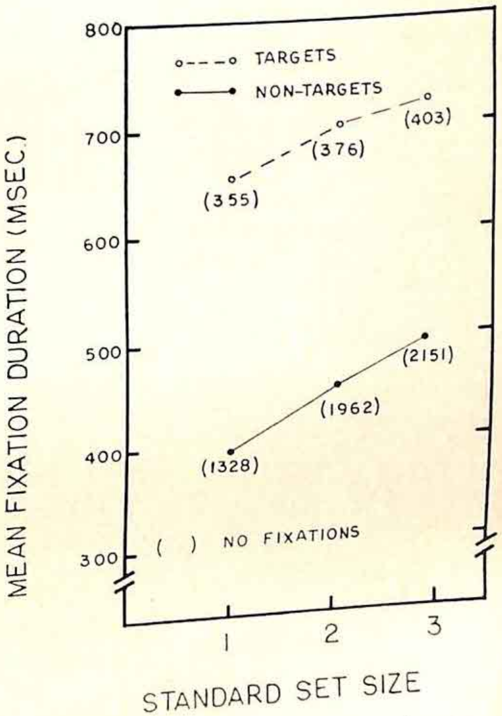


FIG. 9. Fixation durations on target and nontarget characters as a function of standard set size.

mation from noisy items as they do from noiseless items. They compare two pure representations. In a less structured task or in one in which each noiseless item is not fixated individually, such as occurs in real-life visual inspection applications (e.g., Harris, 1966; Schoonard & Gould, 1973), the effect of noise would probably be to reduce the size of the useful field of view and hence increase the number of eye fixations rather than increase the durations of eye fixations.

At microlevel, the present study extends Sternberg's (1967) memory search findings in three ways: (*a*) from a singly presented comparison character to many simultaneously presented ones; (*b*) from noiseless to noisy standards; and (*c*) by providing data on processing times of nontargets as well as on processing times of targets when more than one comparison item is viewed.

Memory search rate, as defined by Sternberg (1967) and as measured by fixation durations on comparison characters, is about 30–50 msec. per item. It is not affected much by the number of comparison items that *S*s have previously fixated. This finding might be predicted by previous results on memory search. It is doubtful, however, that the other main causes of increased overall search times with an increase in memory load, namely the findings of more refixations on both standards and comparison set items, as well as a greater probability of *S*'s scanning past the target character, would have been predicted. Perhaps one reason Briggs and Blaha (1969), who measured overall reaction time, found an apparent increase in memory search rate when the number of displayed items increased 1–4 was because in some cases *S*s made more than one eye fixation on multiple-item displays and in other cases they did not.

The finding that memory search rate, based upon overall reaction times, is constant at about 100 msec. per item (Figure 3), whereas the eye scanning behavior changes with increases in memory set size suggests two different possibilities. On the one hand, the processes underlying memory search might be different depending upon memory set size. This, of course, is inconsistent with Sternberg's (1967, 1969) theory. On the other hand, the processes are the same but the more frequent eye fixations that were found when the standard set was three merely indicates that *S* did not have the standard set well memorized. That is, he sometimes scanned past the target, and he made more refixations on both the standard set and the comparison set.

The linear relationship between fixation duration on a character and its vertical separation from the previously fixated character indicates that the amount of information that *S*s encode about a character fixated peripherally is proportional to its eccentricity. This peripheral encoding is probably gross, perhaps aiding the positional accuracy of the next eye fixation or reducing the number of geometric features that must be subsequently encoded foveally. Yet it is powerful enough to modify encoding times by about 100 msec. over a range of 7°. There is a perfect, negative, rank-order correlation between the vertical and horizontal separations of two adjacent characters, and thus it seems that the vertical separation is dominant over the horizontal one. This can be accounted for by the Chaiken, Corbin, and Volkman (1962) finding that the useful field of view is not as great in the vertical direction as in the horizontal direction.

Fixation durations on target characters were about 250 msec. longer than on nontarget characters. Since the target character was usually the last one fixated on a stimulus, much of this difference might be attributed to response preparation and response bias because, on the average, there were about five times more nontargets present. However, monotonic increasing functions relating the duration of an eye fixation (Gould & Dill, 1969) or overall reaction time (Nickerson, 1967) to the number of geometric features that two patterns have in common have been found. These results suggest a self-terminating feature-comparison process. (Nickerson, however, found that same judgments were faster than different judgments for pairs of stimuli that differed on just one dimension.)

On the other hand, the fixation duration data of Figure 8 are consistent with Sternberg's (1967, 1969) exhaustive memory search interpretation. Thus, two comparison processes have just been described: a self-terminating *feature* or "within" comparison process of the features in a fixated pattern with those in a memory set pattern, and a *unit* or "between" exhaustive comparison process of the fixated item with each memory set item. The relationship between these two processes, although logically simple, is by no means behaviorally clear. Perhaps each applies to a different class of stimuli (cf. Bindra, Donderi, & Nishisato, 1968). Perhaps they are both ongoing, but no present model adequately accounts for them. Recently, Dumas (1972) has also noted unit- and feature-comparison processes.

The present study is not relevant to an issue that several recent experiments have addressed, namely Sternberg's (1967) serial exhaustive interpretation of processing several items at once vs. Neisser's (Neisser et al., 1963) parallel processing findings. The present study is not germane because under the conditions of (*a*) little practice, (*b*) different standard sets on each trial, and (*c*) very low error rates, people cannot process several items as fast as one item (Metlay, Sokoloff, & Kaplan, 1970; Wattenbarger, 1969).

The number of refixations that occurred (less than one per stimulus) suggest, in the absence of data, that *S*s probably do not refixate very frequently in search tasks in which they are instructed not to look back (e.g., Neisser et al., 1963).

REFERENCES

Atkinson, R. C., Holmgren, J. E., & Juola, J. F. Processing time as influenced by the number of elements in a visual display. *Perception & Psychophysics*, 1969, **6**, 321–326.

Bartz, A. E. Fixation errors in eye movements to peripheral vision. *Journal of Experimental Psychology*, 1967, **75**, 444–446.

Bindra, D., Donderi, D., & Nishisato, S. Decision latencies of "same" and "different" judgments. *Perception & Psychophysics*, 1968, **3**, 121–130.

Briggs, G. A., & Blaha, J. Memory retrieval and central comparison times in information processing. *Journal of Experimental Psychology*, 1969, **79**, 395–402.

Chaiken, J. D., Corbin, H. H., & Volkman, J. Mapping a field of short-time visual search. *Science*, 1962, **138**, 1327–1328.

Chase, W. G. Parameters of visual and memory search. Unpublished doctoral dissertation, University of Wisconsin, 1969.

Dumas, J. Scanning memory for multidimensional stimuli with extended practice. *Perception & Psychophysics*, 1972, **11**, 209–212.

Feinberg, R. A study of some aspects of peripheral acuity. *American Journal of Optometry & Archives of American Academy of Optometry*, 1949, **26**, 105–119.

Gould, J. D. Eye movements during visual search. In H. W. Leibowitz (Ed.), *Proceedings of 1969 NATO Symposium on Image Evaluation*, Munich, Germany, 1969.

Gould, J. D., & Dill, A. Eye-movement parameters and pattern recognition. *Perception & Psychophysics*, 1969, **6**, 311–320.

Gould, J. D., & Schaffer, A. Eye movement patterns in scanning numeric displays. *Perceptual Motor Skills*, 1965, **20**, 521–535.

Harris, D. Effect of equipment complexity on inspection accuracy. *Journal of Applied Psychology*, 1966, **50**, 236–237.

Jastrow, J. Studies from the laboratory of experimental psychology of the University of Wisconsin —II. *American Journal of Psychology*, 1892, **4**, 381–428.

Kaplan, I. T., Carvellas, T., & Metlay, W. Visual search and immediate memory. *Journal of Experimental Psychology*, 1966, **71**, 488–493.

Llewellyn-Thomas, E., & Lansdown, E. L. Visual search patterns of radiologists in training. *Radiology*, 1963, **81**, 288–292.

Mackworth, J. F., & Mackworth, N. H. Eye fixations recorded on changing visual scenes by the television eye-marker. *Journal of the Optical Society of America*, 1958, **48**, 439–445.

Mackworth, N. H. Visual noise causes tunnel vision. *Psychonomic Science*, 1965, **3**, 67–68.

Mandelbaum, J., & Sloan, L. L. Peripheral visual acuity. *American Journal of Opthalmology*, 1947, **30**, 581–588.

Metlay, W., Sokoloff, M., & Kaplan, I. T. Visual search for multiple targets. *Journal of Experimental Psychology*, 1970, **85**, 148–150.

Neisser, U., Novick, R., & Lazar, R. Searching for ten targets simultaneously. *Perceptual Motor Skills*, 1963, **17**, 955–961.

Nickerson, R. Response time with a memory-dependent decision task. *Journal of Experimental Psychology*, 1966, **72**, 761–769.

Nickerson, R. "Same"–"different" response times with multi-attribute stimulus differences. *Perceptual Motor Skills*, 1967, **24**, 543–554.

Schoonard, J. W., & Gould, J. D. Field of view and target uncertainty in visual search and inspection. *Human Factors*, 1973, **15**, 33–42.

Sternberg, S. Two operations in character-recognition: Some evidence from reaction-time measurements. *Perception & Psychophysics*, 1967, **2**, 45–53.

Sternberg, S. Memory-scanning: Mental processes revealed by reaction-time experiments. *American Scientist*, 1969, **57**, 421–457.

Sternberg, S., & Scarborough, D. L. Parallel testing of stimuli in visual search. Paper presented at the International Symposium on Information Processing and Control of Motor Activity, Sofia, Bulgaria, July, 1969.

Wattenbarger, B. L. Speed and accuracy set in visual performance. Paper presented at the meeting of the Midwestern Psychological Association, Chicago, May, 1969.

Williams, L. G. The effects of target specification on objects fixated during visual search. *Acta Psychologica*, 1967, **27**, 355–360.

(Received June 22, 1972)

Journal of Experimental Psychology
1973, Vol. 98, No. 1, 196–202

A TWO-PROCESS THEORY FOR THE SHORT-TERM RETENTION OF MOTOR RESPONSES[1]

JOHN L. CRAFT[2]

Wartburg College

A formal model of short-term motor memory was tested. On every trial, *S* executed one linear movement of a constant, standard length. An interfering movement was executed either prior to or following the standard. A variable time interval was interpolated either between the execution of both movements or following both movements. The *S*s were informed which movement was to be recalled either after both had been executed (Experiment I) or before either movement was executed (Experiment II). Results indicated that the interfering effect of one movement upon the recall of another occurs at the time of completion of the second movement, rather than at recall, and that the relative degree of interference is determined by the amount of elapsed time between memory trace establishment and trace interference. Retroactive interference was greater than proactive interference and precuing and postcuing of the recall requirement did not produce differential magnitudes or patterns of response at recall.

The present study was directed at the development and empirical test of a theoretical formulation concerned with the short-term memory (STM) of simple motor responses. The present research attempted to incorporate two mechanisms, trace decay and interference, into a two-process theory of motor STM. The model presented is conceptually related to a two-process theory of motor STM advanced by Pepper and Herman (1970), but is somewhat more formal.

Previous studies have shown that movements or responses executed either prior to or following the execution of a single criterion response reduce the recall accuracy of the criterion response (Stelmach, 1969a; Williams, Beaver, Spence, & Rundell, 1969). Studies by Craft and Hinrichs (1971) and Herman and Bailey (1970) further demonstrated that retroactive interference (RI) effects are greater than proactive interference (PI) effects.

Interference effects in motor STM have also been shown to be a function of the specific similarity between motor responses. Pepper and Herman (1970) found that interpolated forces, greater in magnitude than a criterion force, produced recall of larger forces than did interpolated forces smaller than the criterion. Herman and Bailey (1970) replicated this finding for interpolated forces, but failed to observe a similar effect when greater or lesser magnitude forces preceded the criterion force. Stelmach (1969b) varied the absolute difference between four prior lever responses and a criterion response and found a direct relationship between mean absolute difference and recall error of the criterion. However, directional error trends as a function of response similarity could not be assessed because the prior movements were a mixture of responses both shorter than and longer than the criterion. Craft and Hinrichs (1971) did observe directional error trends as a function of the similarity between two motor responses. Their data indicated that as the similarity between a standard and an interfering movement decreased, error at recall of the standard movement increased in the direction of the length of the interfering movement.

---

[1] This research was submitted in partial fulfillment of the requirements for the PhD degree at the University of Iowa. This research was supported in part by U. S. Public Health Service Research Grant MH-16362 from the National Institute of Mental Health. Appreciation is expressed to the members of the dissertation committee: cochairmen J. Richard Simon and James V. Hinrichs for their individual and combined efforts during the development of this research, and Irwin P. Levin, John R. Platt, and Leon Smith for their many helpful comments and criticisms.

[2] Requests for reprints should be sent to John L. Craft, Department of Psychology, Wartburg College, Waverly, Iowa 50677.

Pepper and Herman (1970) incorporated both trace decay and interference features into their model to account for RI effects. They hypothesize that interference occurs during the retention interval in the form of an alteration of the original memory trace. The present model assumes a similar interference mechanism and is designed to account for both RI and PI effects. The major aspects of the present model are contained in the following set of assumptions.

An accurate memory trace or internal representation of the intensity of a motor response is initially stored, but is subject to decay over time. Assuming a one to one correspondence between trace intensity and representation of response extent, the decay occurs along the dimension of intensity of the response in the direction of lesser extent. The memory traces, T_1 and T_2, produced by two movements, M_1 and M_2, are assumed to decay independently as an exponential function of time.

Two alternative loci of interference can be assumed. In one case it can be assumed that the traces produced by the two movements will both be available at the moment of recall. In this case, interference effects of M_1 on recall of M_2, or of M_2 on recall of M_1, occur at recall in the form of memory trace alteration. In the other case, interference effects of M_1 on recall of M_2, or of M_2 on recall of M_1, are assumed to occur following the completion of M_2 rather than at the moment of recall.

The degree of interference of one movement upon recall of another will be determined by the relative weighting of the two trace intensities at the moment of trace alteration. In a general form, the magnitude of response (L) at recall can be expressed as

$$L = \frac{\Sigma w_j T_j}{\Sigma w_j},$$

where T_j represents the intensity of the M traces at recall and w_j represents the relative weighting of the traces. The present research is mainly concerned with the effects of the variable time (t) on interference effects at recall. Other potential factors that might affect the weightings are assumed to remain constant across variations in t. Two alternative forms of the weighting assumption can be stated. In one case, the w_js are assumed to be independent of time. In the other case, the w_js are assumed to be an inverse function of time.

From this general model for L, three specific models can be derived by combining the two alternative loci of interference assumptions and the two alternative trace-weighting assumptions. Model 1 assumes that the w_js are independent of time and that interference occurs at the moment of recall (R). This model suggests the magnitude of L at recall will be a simple weighted average of the trace intensities at the moment of trace alteration, independent of the locus of the trace alteration (at M_2 vs. R). Model 2 assumes that the w_js are inversely related to time and that interference occurs at the moment of recall. Since interference is assumed to occur at recall for this model, both the $M_1 - R$ and $M_2 - R$ intervals are relevant factors in determining the relative weightings. Model 3 assumes that interference occurs at M_2 and that w_1 is the only weighting which is a function of time. This model predicts that the magnitude of L at recall will be a weighted average of the trace intensities, but the only relevant time interval determining the weightings is the $M_1 - M_2$ interval.

The present study was designed to test these three alternative models. The *S*s' task was to execute two blind movements of controlled length and to recall one of the movements. On every trial, *S* executed one movement of a constant, standard (S) length. The other movement, defined as an interfering (I) movement, was executed either prior to or following the standard. On each trial, *S* attempted to recall the standard movement. Variables manipulated were (*a*) the length of the interfering movement, either the same, shorter, or longer than the standard; (*b*) the locus of the interfering movement, either prior to or following the standard; and (*c*) the locus of a variable time interval (Δt), either following the execution of both movements or interpolated between the execution of the two movements.'

Separate sets of predictions were made for each of the two loci of interfering movement conditions. Those experimental conditions in which the interfering movement follows the standard are referred to as poststandard conditions, and those in which the interfering movement precedes the standard are referred to as prestandard conditions.

The major interest in the present study was to determine how the slope of the function relating L_S (recall of the standard) to length of I changes as a function of the time manipulation, Δt. The slope of the function relating L_S to length of I is interpreted as reflecting the magnitude of interference produced by I at the moment of trace alteration on the recall of S. For each of the two loci of I conditions (poststandard and prestandard), specific predictions of the relative slope of this function were made based on the assumptions of each model.[3] For each model, the slope of the L_S function was derived from the value of the L_S function at the moment of trace alteration.

The poststandard conditions can now be represented as three distinct types of trials, S-I-R, S-I-Δt-R, and S-Δt-I-R, or in simpler notation, A, B, and C. Model 1 predicts that the order of the slopes of the function relating L_S to I will be $A = C > B$; Model 2, $C > A > B$; and Model 3, $C > A = B$. The prestandard conditions can be represented as I-S-R, I-S-Δt-R, and I-Δt-S-R trials, or D, E, and F. Model 1 predicts that the ordering of the slopes will be $D > E = F$; Model 2, $E > D > F$; and Model 3, $D = E > F$.

In summary, the major interest in the present study was to determine how the slope of the function relating L_S to length of I changes as a function of the time manipulation, Δt. The critical test of a difference in slope of the L_S function for any two conditions was the significance of the interaction between those two conditions and length of I. A prediction of equality of slope for any two conditions was taken to mean that the interaction between these two conditions and length of I would not be statistically significant. The observation of specific interactions and the specific pattern of slope magnitudes within each of the two loci of I conditions determined the degree of support for each of the separate sets of assumptions incorporated into the three models.

METHOD

Subjects. The *S*s were 84 right-handed females enrolled in the introductory psychology courses at the University of Iowa.

Apparatus. The apparatus consisted of two free-moving 5-cm. square Plexiglas blocks mounted on a wooden meter stick. One block could be locked in place by *E* to define the length of the standard or interfering length. The other block had a knob for fingertip grasping and was moved by *S* to present the standard or interfering lengths or to indicate *S*'s response at recall. The apparatus was placed on a table parallel to the coronal plane, with the direction of *S*'s movement being from left to right. The *S* was blindfolded to ensure that her responding was not based on visual cues.

Experimental design. The variables manipulated (all within *S*) were: paradigm (S-I-R or I-S-R), length of I (24, 34, 44, 54, and 64 cm.), and interpolated interval (0 sec. and 20 sec. between M_1 and M_2, and 20 sec. between M_2 and R). Thus, the six paradigm–interval combinations were S-I-R, S-20-I-R, S-I-20-R, I-S-R, I-20-S-R, and I-S-20-R. Each *S* performed on one block of 30 trials. For each of the six paradigm–interval combinations, a standard length of 44 cm. was paired with each of the five lengths of I. On each trial, *S* was asked to repeat the length of the standard. The 42 *S*s in Experiment I were not informed until the time of recall which movement was to be repeated. The 42 *S*s in Experiment II were informed both prior to and following the movements which one was to be repeated. Furthermore, *S* was not informed that the same length (the standard) would be presented on each trial. The assignment of length of I to each successive occurrence of a particular paradigm–interval combination was random for each *S*.

Procedure. On all trials, *S* heard the following tape-recorded sequence of commands: "engage," "move one," "return," "move two," "return," "repeat" (either "one" or "two"), and "return and disengage." Upon hearing the command "engage," *S* grasped the knob and waited for the first "move" command. At the "move" command, *S* was instructed to "smoothly move your block down the slide, to your right, until your movement is stopped by the other block." The *S* held her block at the stop until the command "return" was presented. The *S* then returned her block to the starting position. Following the second "move" and "return" commands, *S* heard either "repeat one" or "repeat two" which "will be a signal to repeat the length of your first or second movement, respectively." At the "return and disengage" command, *S* returned her block and removed her hand from the apparatus until the next "engage" command. The interval between the "disengage" command on one trial and the "engage" command on the next was 15 sec. For S-I-R and I-S-R trials, the interval between each command within a sequence was 5 sec. On I-S-20-R and S-I-20-R trials, the interval between the "return" command following the second movement and the "repeat" command was increased 5–25 sec. On I-20-S-R and S-20-I-R trials, the interval between the "return" command following the first movement and the "move two" command was increased 5–25 sec. The *S* was not permitted to remove her hand from the apparatus during a trial. Four practice trials were given prior to the test block of trials.

---

[3] The detailed equations and mathematical slope derivations for each model are given in Craft (1971).

RESULTS

Since the predictions made from each model presupposed a linear relationship between L_S and length of I, the data obtained for each of the six conditions were subjected to trend analyses testing linear, quadratic, and cubic components. For each condition, the trend analyses involved the use of orthogonal polynomials to obtain linear, quadratic, and cubic comparison values for individual *S*'s data. The mean of these comparison values was tested against the value of zero for each of the three trends. The linear trend component was significant ($p < .01$) for all conditions except I-20-S-R, and all other trends were nonsignificant.

Since predictions from the models were made separately for the poststandard and prestandard conditions, analyses of variance were carried out separately for these two loci of I conditions. Magnitude of response at recall of the standard (L_S) as a function of length of I for each of the three poststandard conditions in Experiment I is presented in the top half of Figure 1. Figure 1 shows that L_S increases as length of I increases for each of the three poststandard conditions. Separate analyses of variance indicated that length of I was a significant effect ($p < .025$) for each condition.

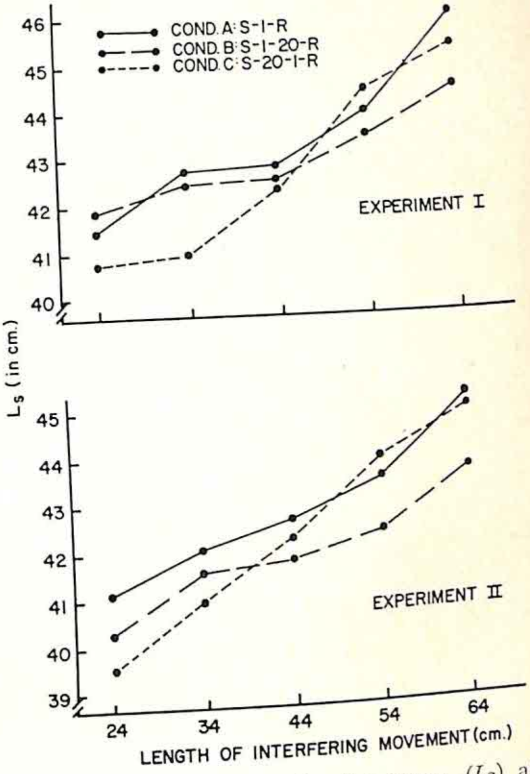


FIG. 1. Mean magnitude of response (L_S) at recall of the standard for the poststandard conditions of Experiments I and II.

An analysis involving only S-20-I-R and S-I-20-R trials revealed a significant ($p < .05$) Length of I × Interval interaction. Figure 1 shows that the slope of the function relating L_S to length of I is greater for S-20-I-R than for S-I-20-R trials. Although the slope of this function appears to be greater for S-20-I-R than for S-I-R trials, this apparent interaction was not borne out by statistical analysis ($p > .05$). Similarly, when only the S-I-R and S-I-20-R trials were subjected to analysis, a significant Length of I × Interval interaction was not observed. Though the only clear statistical difference in slopes is S-20-I-R > S-I-20-R, in terms of trends, the ordering of the slopes appears to be slope for S-20-I-R > S-I-R = S-I-20-R; or in simpler notation, C > A = B.

Magnitude of response at recall of the standard (L_S) for each of the three prestandard conditions in Experiment I is presented in the top half of Figure 2. Figure 2 shows that L_S increases as length of I increases for the I-S-R and I-S-20-R conditions, but not for the I-20-S-R condition.

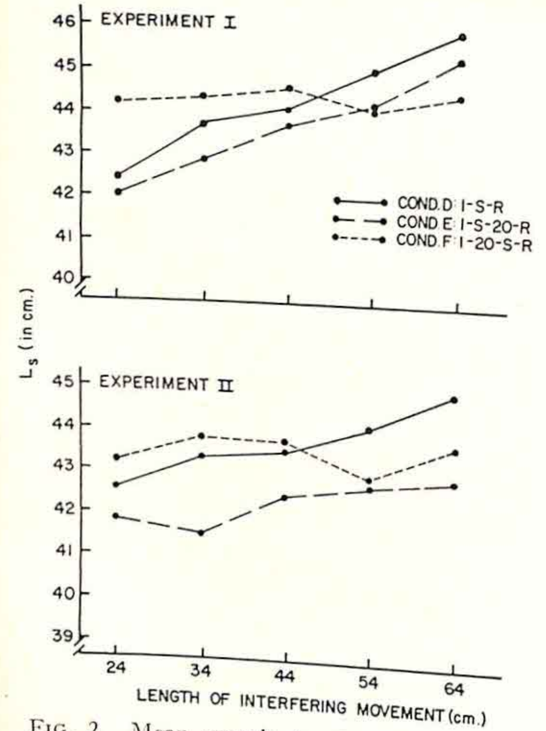


FIG. 2 Mean magnitude of response (L_S) at recall of the standard for the prestandard conditions of Experiments I and II.

Separate analyses of variance revealed that length of I was a significant source of variance for the I-S-R and I-S-20-R conditions ($p < .001$), but not for the I-20-S-R condition.

An analysis involving only I-S-R and I-20-S-R trials revealed a significant Length of I × Interval interaction ($p < .001$). Figure 2 shows that the slope of the function relating L_S to length of I is greater for I-S-R than for I-20-S-R trials. When only I-S-20-R and I-20-S-R trials were subjected to analysis, a significant Length of I × Interval interaction was also observed ($p < .001$). Note in Figure 2 that the slope of the function relating L_S to length of I is greater for I-S-20-R trials than for I-20-S-R trials. This latter analysis also indicated that mean L_S was greater for I-20-S-R trials than for I-S-20-R trials (44.3 vs. 43.6 cm., $p < .05$). An analysis involving only I-S-R and I-S-20-R trials indicated that the Length of I × Interval interaction was not significant. Mean L_S was greater on I-S-R than on I-S-20-R trials (44.2 vs. 43.6 cm., $p < .05$). Based on the outcome of the statistical analyses as well as observed trends, then, the ordering of the slopes for the prestandard conditions is slope of L_S for I-S-R = I-S-20-R > I-20-S-R. Using the simpler notation, D = E > F.

Although no predictions were made from the models with respect to differences across locus of I conditions, an overall analysis was carried out. A significant Locus of I × Length of I interaction ($p < .05$) indicated that the slope of the function relating L_S to length of I is greater for the poststandard conditions than for the prestandard conditions. In addition, L_S was significantly ($p < .001$) greater for the prestandard than for the poststandard conditions (44.1 vs. 43.0 cm.).

Results of Experiment I suggest that Model 3 is the most appropriate of the three alternative models for the present experimental situation. The predictions generated from Model 3 with respect to the ordering of slope magnitudes were clearly confirmed in the prestandard conditions, but only partially confirmed in the poststandard conditions.

Experiment II was a replication of Experiment I, with the added feature that *S*s were precued as to the response requirement at recall. Craft and Hinrichs (1971) observed that response variance was less when *S*s were precued than when they were postcued. If precuing is effective in reducing response variance without producing differences in response pattern or magnitude, then the apparent difference in slope for S-20-I-R and S-I-R conditions shown in Figure 1 might also be borne out by statistical analysis, and thus provide unequivocal support for Model 3.

The results of Experiment II are displayed in the bottom halves of Figures 1 and 2. Except for a few minor discrepancies, response magnitude and pattern were essentially the same in the two experiments. In terms of the interactions observed, the results exactly paralleled those of Experiment I. However, precuing of the response requirement at recall in Experiment II was not effective in reducing the response variance below the level observed in Experiment I. Hence, the apparent difference in slope of the L_S function for S-20-I-R and S-I-R trials shown in Figure 1 still eluded statistical confirmation, suggesting that the failure to confirm this difference might be a problem of insufficient statistical power. Therefore, although Experiments I and II were conducted as separate experiments, it was decided that an analysis of the combined data of the two experiments would be informative. An analysis of the combined data involving only the S-20-I-R and S-I-R conditions now revealed a significant ($p <$ .025) Length of I × Interval interaction. Taken together, the results of the combined data analyses within both the poststandard and prestandard conditions offer unequivocable support for the predictions made based on the assumptions of Model 3.

DISCUSSION

The results of the present study provide evidence that Model 3 is the most appropriate of the three alternative models as a conceptual framework for the present experimental situation. Results indicate that the interfering effect of one movement upon the recall of another occurs at the time of completion of the second movement, rather than at recall. Furthermore, the relative degree of interference is determined by the amount of elapsed time between trace establishment and trace alteration. These findings suggest that the only relevant time interval in determining the relative degree of interference of one movement upon recall of another is the interval between the two movements. The time interval between the second movement and recall of the standard appears only to affect the overall magnitude of response at recall, with magnitude decreasing as the interval increases.

The poststandard conditions in the present study can be considered as a set of RI conditions. Similarly, the prestandard conditions can be considered as a set of PI conditions. Using these terms, then, the results of the present study can be interpreted as RI increasing and PI decreasing when a Δt is interpolated between M_1 and M_2, and no increase in PI or decrease in RI when a Δt is interpolated between M_2 and R. Underwood and Ekstrand (1967) have demonstrated that PI decreases when successive verbal tasks are separated by a time interval and attribute the reduction of PI to an increase in list differentiation. However, the finding that an increase in the *M_2 — R interval did not produce a corresponding increase in PI or decrease in RI is* contrary to one of the hypotheses of classical interference theory for the long-term retention of verbal materials. According to classical interference theory (e.g., Keppel, 1968), PI should increase and RI should decrease as the retention interval increases. This hypothesis is based on the assumption that interference occurs at recall, rather than during the retention interval, and is similar to Model 2 presented in the present study.

Herman and Bailey (1970) concluded, on the basis of their findings, that while RI effects in motor STM are readily apparent, PI effects in motor STM are, at best, weak. The present study, however, demonstrated that an interfering movement may act both proactively and retroactively in producing interference with recall of a standard movement, but that RI is a more potent source of interference with a motor response than is PI. In other words, the magnitude of interference produced by one motor response upon recall of another is dependent upon the sequential order of the two responses. This finding is also in empirical agreement with studies of verbal STM, where it has been observed that PI effects are of lesser magnitude than are RI effects (e.g., Postman, 1964, p. 160; Wickelgren, 1966).

Results of the present study also replicate previous findings that response interference effects in motor STM follow a dimension of length similarity between motor responses (Craft & Hinrichs, 1971; Patrick, 1971). It was observed in the present study that the longer or shorter the interfering length, the greater the change in response magnitude toward the length of the interfering movement. This directional biasing effect may be related conceptually to the classical data on time error, which indicated that a standard trace is "assimilated" to or toward an interpolated stimulus (Woodworth & Schlosberg, 1954, p. 229). The inverse relationship between similarity and interference effects for motor STM is at variance with findings in verbal STM, where interference has been shown to be a direct function of the similarity of verbal materials (e.g., Kintsch, 1970).

Another important finding of the present study was that there were no major differences in the results of Experiments I and II. The remarkable similarity in both pattern and magnitude of response across experiments supports and extends the similar finding of Craft and Hinrichs (1971) that the locus of cuing of the required response at recall does not appear to be a relevant variable in motor STM. This lack of difference between precuing and postcuing of the required response is also at variance with results of studies of verbal STM, where precuing of the response order required at recall typically leads to better recall performance than does postcuing (e.g., Hinrichs, 1968).

REFERENCES

Craft, J. L. Memory trace interference and decay as factors influencing the short-term retention of motor responses. Unpublished doctoral dissertation, University of Iowa, 1971.

Craft, J. L., & Hinrichs, J. V. Short-term retention of simple motor responses: Similarity of prior and succeeding responses. *Journal of Experimental Psychology*, 1971, **87**, 297–302.

Herman, L. M. & Bailey, D. R. Comparative effects of retroactive and proactive interference in motor short-term memory. *Journal of Experimental Psychology*, 1970, **86**, 407–415.

Hinrichs, J. V. Prestimulus and poststimulus cuing of recall order in the memory span. *Psychonomic Science*, 1968, **12**, 261–262.

Keppel, G. Retroactive and proactive inhibition. In T. R. Dixon & D. L. Horton (Eds.), *Verbal behavior and general behavior theory*. Englewood Cliffs, N. J.: Prentice-Hall, 1968.

Kintsch, W. *Learning, memory, and conceptual processes*. New York: Wiley, 1970.

Patrick, J. The effect of interpolated motor activities in short-term motor memory. *Journal of Motor Behavior*, 1971, **3**, 39–48.

Pepper, R. L., & Herman, L. M. Decay and interference effects in the short-term retention of a discrete motor act. *Journal of Experimental Psychology Monograph*, 1970, **83**(2, Pt. 2).

Postman, L. Short-term memory and incidental learning. In A. W. Melton (Ed.), *Categories of human learning*. New York: Academic Press, 1964.

Stelmach, G. E. Prior positioning responses as a factor in short-term retention of a simple motor task. *Journal of Experimental Psychology*, 1969, **81**, 523–526. (a)

Stelmach, G. E. Short-term motor retention as a function of response similarity. *Journal of Motor Behavior*, 1969, **1**, 37–45. (b)

Underwood, B. J., & Ekstrand, B. R. Studies of distributed practice: XXIV. Differentiation and proactive inhibition. *Journal of Experimental Psychology*, 1967, **74**, 574–580.

Wickelgren, W. A. Phonemic similarity and interference in short-term memory for single letters. *Journal of Experimental Psychology*, 1966, **71**, 396–404.

Williams, H. L., Beaver, W. S., Spence, M. T., & Rundell, O. H. Digital and kinesthetic memory with interpolated information processing. *Journal of Experimental Psychology*, 1969, **80**, 530–536

Woodworth, R. S., & Schlosberg, H. *Experimental psychology*. New York: Holt, 1954.

(Received June 29, 1972)

Journal of Experimental Psychology
1973, Vol. 98, No. 1, 203–205

SHORT REPORTS

MEMORY FOR TACIT IMPLICATIONS OF SENTENCES[1]

MARCIA K. JOHNSON,[2] JOHN D. BRANSFORD, AND SUSAN K. SOLOMON
State University of New York at Stony Brook

The present study investigated the nature of the information that is available to *Ss* when they have comprehended linguistic materials. The *Ss* were read a set of acquisition stories and later asked to recognize which of a set of test sentences they had actually heard in the stories. The *Ss* were likely to "recognize" novel sentences which expressed information consonant with implications of the material presented during acquisition.

The present experiment was based on the assumption that comprehension involves inferential thinking. For example, a study by Bransford, Barclay, and Franks (1971) indicated that information about spatial relationships implied but not explicitly stated in linguistic inputs is available to *Ss* on a later recognition task. However, comprehending many types of information may include more than knowledge about the spatial relations among objects. For example, *Ss* may also make inferences about the objects which might be involved in certain relations, and about the consequences suggested by certain relations. The present study asked whether *Ss* were likely to think they heard information based on such inferences.

Two classes of items were investigated. Type A items were like those previously studied by Kintsch (1972). Kintsch presented *Ss* with sentences like THE MAN WAS SHOT and asked them to specify additional information that seemed to be true about the sentences. His *Ss* indicated that the man must be shot by something (i.e., there must be some *instrument* for carrying out the action, to use Fillmore's, 1968, term), and most assumed that "something" would be a gun. Type B items each suggested some consequence of the action described, e.g., THE BOY HIT THE BASEBALL AND WATCHED AS IT FLEW INTO THE PICTURE WINDOW IN THE HOUSE. One probable consequence is that the window shattered. The *Ss* heard stories of both types and were then presented with a recognition test where some items contained information that was only true by implication (e.g., that a gun was used to shoot the man in Type A situations, that the window shattered in Type B situations).

Method. The experiment consisted of an acquisition phase and a recognition phase. Experimental and control groups received the same recognition list and differed with respect to the acquisition list.

For both groups, the acquisition list consisted of 20 short (two or three sentences) descriptive stories. Eight of these were identical in both experimental and control lists and were filler items which were of no interest in the analysis. For the experimental group, the remaining 12 stories were inference stories. There were 6 Type A stories in which an object was implied and 6 Type B stories, each of which suggested some probable consequence. For the control group, the same 12 inference story frames were used, but in each case either a prepositional phrase or a verb phrase was changed. Thus, for the control group no object was implied or the implied object was different in Type A stories, and the probable consequence of the action was changed in Type B stories.

The recognition list consisted of 38 sentences. The first 2 sentences were filler sentences included to give *Ss* an opportunity to adjust to the recognition task and to get used to the pacing. The remaining 36 sentences included three categories—old, unrelated, and inference. Old sentences were identical to those used in the stories of both groups. Unrelated sentences involved elements from the acquisition stories, but were inconsistent with the information conveyed by the stories. Inference sentences were not identical to the acquisition sentences for either experimental or control groups, but were potentially true by implication for experimental *Ss*.

Three examples of each type of acquisition story are given below, with the changes made for the control group given in parentheses and the corresponding inference recognition item given in italics.

Type A: (*a*) JOHN WAS TRYING TO FIX THE BIRD HOUSE. HE WAS POUNDING (LOOKING FOR) THE NAIL WHEN HIS FATHER CAME OUT TO WATCH HIM AND TO HELP HIM DO THE WORK. *John was using the hammer to fix the bird house when his father came out to watch him and to help him do the work.* (*b*) THE FARMERS MUST BE WARNED OF THE ONCOMING FLOOD, THE SHERIFF CRIED. HE MOUNTED (PHONED) AS QUICKLY AS POSSIBLE SINCE HE KNEW THAT IT WOULD TAKE QUITE A WHILE TO SPREAD ALL THE NEWS. *The sheriff mounted his horse as quickly as possible since he knew that it would take quite a while to spread all the news.* (*c*) IT WAS 7:00 ON MONDAY MORNING AND THE MAN DROVE (WALKED) TO WORK. IT WAS A BEAUTIFUL DAY, AND HE REALLY ENJOYED THE TRIP. *It was 7:00 on Monday morning and the man took his car to work.*

---

[1] This research was supported in part by a Research Foundation of the State University of New York summer research fellowship to the first author. The authors wish to thank Brian O'Callaghan, guidance counselor, and the students of Ward Melville High School, East Setauket, New York, for their cooperation. Senior authorship was decided between the first two authors by flipping a coin.

[2] Requests for reprints should be sent to Marcia K. Johnson, Department of Psychology, State University of New York, Stony Brook, New York 11790.

TABLE 1

Mean Number of Yes Responses during Recognition

Group	Recognition item	
	Old (10 sentences)	Inference (12 sentences)
Experimental	6.65	7.45
Control	6.95	2.60

Type B: (*a*) WHEN THE MAN ENTERED THE KITCHEN HE SLIPPED ON A WET SPOT AND DROPPED (JUST MISSED) THE DELICATE GLASS PITCHER ON THE FLOOR. THE PITCHER WAS VERY EXPENSIVE, AND EVERYONE WATCHED THE EVENT WITH HORROR. *When the man entered the kitchen he slipped on a wet spot and broke the delicate glass pitcher when it fell on the floor.* (*b*) IT WAS LATE AT NIGHT WHEN THE PHONE RANG AND A VOICE GAVE A FRANTIC CRY. THE SPY THREW (PULLED) THE SECRET DOCUMENT INTO (FROM) THE FIREPLACE JUST IN TIME, SINCE 30 SECONDS LONGER WOULD HAVE BEEN TOO LATE. *The spy burned the secret document just in time since 30 seconds longer would have been too late.* (*c*) THE RIVER WAS NARROW. A BEAVER HIT THE LOG THAT A TURTLE WAS SITTING ON (BESIDE) AND THE LOG FLIPPED OVER FROM THE SHOCK. THE TURTLE WAS VERY SURPRISED BY THE EVENT. *A beaver hit the log and knocked the turtle into the water.*

For both experimental and control *S*s the same ordering of the 20 acquisition stories was used. The order was random with the restriction that the acquisition list begin and end with a filler story. The recognition sentences were randomized in blocks of the three categories (old, unrelated, and inference). Due to an error in making up the control acquisition list, two recognition sentences which should have been old for these *S*s were not (one each from Type A and Type B stories). Therefore, the corresponding recognition sentences in the experimental group were ignored when the data were scored. Consequently, for both experimental and control groups, the analysis of the recognition performance was based on 10 old sentences, 12 unrelated sentences, and 12 inference sentences.

Experimental *S*s were tested in two group sessions, as were control *S*s. The *S*s were told to listen carefully to the descriptive stories and that they would be asked questions about them later. The acquisition list was read once by *E* at a normal reading rate. There were approximately 2 sec. between stories, during which *E* announced "next story." The retention interval was 3 min. During this time, paper was distributed, *S*s numbered the page 1–38, and the following recognition instructions were read:

> Now I'm going to read you some sentences. For each one, please indicate whether or not it is an actual sentence from one of the stories you heard. If you think the sentence was exactly as you heard it in one of the stories write yes; if you think the sentence was not from one of the stories, or is not exactly in the same words as you heard it in one of the stories, write no. Please give a yes or no answer for each sentence, even if you feel you have to guess. There will not be much time between sentences, so please make your judgments quickly.

The test sentences were read at a 10-sec. rate.

The experiment was initially conducted with high school student volunteers and was later replicated with introductory psychology students at the State University of New York at Stony Brook. Protocols from 10 *S*s in each of the four testing sessions were randomly selected to equalize *n*s.

Results. In a preliminary analysis of the data, type of *S* (high school or college student) was not a significant variable in any comparison. In addition, the results reported below held for both *S* populations independently. Therefore, the data from both populations were combined in the following analyses.

Experimental and control *S*s did not differ in mean number of yes responses to unrelated items; the means were .70 and .45, respectively ($F < 1$). Table 1 shows the mean number of yes responses for old and inference recognition items. The Condition × Type of Item interaction was significant, $F\ (1,\ 38) = 27.86$, $p < .001$. This interaction reflects the fact that the experimental and control *S*s did not differ on old items ($F < 1$), whereas the experimental *S*s were much more likely than the control *S*s to say yes to the inference items, $F\ (1,\ 38) = 49.73$, $p < .001$. The absolute number of *S*s responding yes was higher in the experimental than control condition for all 12 inference recognition items.

Discussion. These results are consistent with the notion that an *S*'s understanding depends not only on what he hears, but on the implications of this information in light of relevant knowledge he already possesses. The *S*s were likely to think that information available only by implication was actually given during the acquisition task. On the basis of the present procedures one cannot determine whether such implications were actually realized at acquisition or whether they were determined at recognition. The general conditions under which inferences are made is a question for future research. Also of interest are the processes by which *S*s reject information; these may involve acts of inference as well. Consider, for example, the inference test sentence WHEN THE MAN ENTERED THE KITCHEN HE SLIPPED ON A WET SPOT AND BROKE THE DELICATE GLASS PITCHER WHEN IT FELL ON THE FLOOR. Control *S*s could have rejected this sentence because their understanding of the situation included an intact rather than broken pitcher, or because they remembered that the pitcher was on the floor, whereas the verb TO FALL implied that it was initially above the floor. Perhaps the ease (e.g., in a reaction time task) with which an *S* can accept and reject information would provide more precise information about the state of his knowledge following comprehension. The major point of the present article is that slight changes in inputs can result in large differences in the ideas communicated, including the implications that may be derived.

REFERENCES

BRANSFORD, J. D., BARCLAY, J. R., & FRANKS, J. J. Sentence memory: A constructive versus interpretive approach. *Cognitive Psychology*, 1971, 2, 331–350.

FILLMORE, C. J. The case for case. In E. Bach & R. T. Harms (Eds.), *Universals in linguistic theory*. New York: Holt, Rinehart & Winston, 1968.

KINTSCH, W. Notes on the structure of semantic memory. In E. Tulving & W. Donaldson (Eds.), *Organization and memory*. New York: Academic Press, 1972.

(Received November 22, 1971)

Manuscripts Accepted for Publication in the *Journal of Experimental Psychology*

Semantic Processing without Permanent Storage. George Mandler (Department of Psychology, University of California, San Diego, P. O. Box 109, La Jolla, California 92037) and Patricia E. Worden.

Solution of Comparative and Negative-Equative Three-Term Series Problems. Terrence J. Keeney (Department of Psychology, University of California, Riverside, California 92502) and Donald L. Gaudino.

Asymmetries in Processing the Terms "Right" and "Left." Gary M. Olson (Research Laboratory, Naval Submarine Medical Center, Naval Submarine Base New London, Groton, Connecticut 06340) and Kevin Laxar.

Conditioned Adaptation of Prismatic Displacement: Training Trials and Decay. Jerome H. Kravitz (Eye Research Foundation of Bethesda, 8710 Old Georgetown Road, Bethesda, Maryland 20014) and Fred L. Yaffe.

Modified Response Recovery: The Effects of Noncontingent Reward and Nonreward on Spontaneous Recovery. Alfred L. Morin (Department of Psychology, University of Miami, Coral Gables, Florida 33124).

Effect if Prewarning on Increase in Reaction Time in an Auditory Monitoring Task. Hans-Olof Lisper (Department of Psychology, University of Uppsala, Svartbacksgatan 10, S-753 20 Uppsala, Sweden), Lennart Melin, and Per-Olow Sjöden.

Configural Effect in Multiple-Cue Probability Learning. Stephen E. Edgell and N. John Castellan, Jr. (Department of Psychology, Psychology Building, Indiana University, Bloomington, Indiana 47401).

Intentional and Incidental Learning in Children and the von Restorff Effect. F. Michael Rabinowitz (Department of Psychology, Newcomb College, Tulane University, New Orleans, Louisiana 70118) and Susan Ring Andrews.

Intralist Cuing Following Retroactive Inhibition of Well-Learned Items. Milton Blake (Department of Psychology, University of Toronto, Toronto 181, Ontario, Canada) and Ronald Okada.

Recall of Haptic Information by Blind and Sighted Individuals. Joan Shagan and Jacqueline Goodnow (Program in Cognitive and Perceptual Development, George Washington University, Washington, D.C. 20006).

Generalization of Extinction Gradients: A Systematic Analysis. William J. Dubin and Donald J. Levis (State University of New York at Binghamton, Binghamton, New York 13019).

Acquisition and Extinction of the Human Eyelid CR as a Function of the Schedule of Reinforcement and UCS Intensity under Two Masked Conditioning Procedures. Bryce C. Schurr (Department of Psychology, University of Alberta, Edmonton, Alberta, Canada) and Willard N. Runquist.

False Recognition as a Function of Associative Proximity. Janice Vogt (Department of Psychology, Muenzinger Building, University of Colorado at Boulder, Boulder, Colorado 80302) and Gregory Kimble.

Unit-Sequence Interference in Short-Term Memory: Facilitation versus Interference Factors. Anton K. Saba and Thomas W. Turnage (Department of Psychology, Old Botany Hall, Iowa State University, Ames, Iowa 50010).

A U-Shaped Detection Function for Backward Masking of Similar Contours. Sara Burchard and R. B. Lawson (Department of Psychology, John Dewey Hall, University of Vermont, Burlington, Vermont 05401).

Attention and Processing Capacity in Auditory Recognition. Jeffrey J. Moore and Dominic W. Massaro (Department of Psychology, University of Wisconsin, Psychology Building, Charter at West Johnson, Madison, Wisconsin 53706).

Imagery Effects in Continuous Paired-Associate Learning. Edward J. Rowe (Department of Psychology, Memorial University of Newfoundland, St. John's, Newfoundland, Canada) and Shannon K. Smith.

Morality Judgment: Test of an Averaging Model with Differential Weights. Michael H. Birnbaum (Department of Psychology, University of California at Los Angeles, Los Angeles, California 90024).

Retroactive Facilitation in Short-Term Retention of Minimally Learned Paired Associates. Darryl Bruce (Department of Psychology, Florida State University, Tallahassee, Florida 32306) and George E. Weaver.

(*Continued on page 208*)

Journal of Experimental Psychology
1973, Vol. 98, No. 1, 206–208

EFFECT OF INDUCED MUSCLE TENSION ON ACQUISITION AND RETENTION OF VERBAL MATERIAL

HELEN C. BEH[1] AND CAROLE A. HAWKINS

University of Sydney, Sydney, New South Wales, Australia

The effect of induced muscle tension on the acquisition and retention of a serial list of 12 disyllabic words was studied, using four groups of *S*s either learning and recalling under tension, learning under tension and recalling under relaxation, learning under relaxation and recalling under tension, or learning and recalling under relaxation. The *S*s learning under tension were found to require significantly fewer trials to learn the list to a criterion of one errorless trial, although no significant difference was found between *S*s learning under tension and *S*s learning under relaxation in terms of the number of errors made in reaching the learning criterion. No difference was found between tension and relaxation groups in terms of errors during a recall task 10 min. after training, although groups learning under tension demonstrated significantly better recall 1 wk. after training. Tension during recall was found to have no significant effect on recall either 10 min. or 1 wk. after training.

Early studies of the effects of experimentally induced muscle tension (IMT) on the acquisition of verbal material indicate that moderate IMT during training results in increased efficiency of learning, whether such efficiency is measured in terms of trials to criterion (Bills, 1927) or errors after a given number of training trials (Bills, 1927; Courts, 1939; Stauffacher, 1937). A more recent study, however, presents evidence that groups trained during either IMT or muscular relaxation (MR) do not differ significantly in terms of errors made immediately following a single training trial (Bourne, 1955).

Similarly conflicting evidence is found when findings from studies of the effect of differing degrees of muscle tension during training on retention of verbal material are considered. Bills (1927), for example, reported that IMT-trained *S*s demonstrated superior retention when tested 3 hr. after the training period, while Stauffacher (1937) found no difference in retention between IMT- and MR-trained groups after a 24-hr. training–test interval. Bourne (1955) also reported no difference between such groups for training–test intervals of up to 4 min. The study by Bourne, in fact, suggests that the level of muscle tension existing during the recall (test) period affects recall, rather than the level of tension during training, although Stauffacher reported that IMT during recall did not significantly affect 24-hr. recall, and Uehling and Sprinkle (1968) present findings in agreement with those of Stauffacher.

On the whole, the effects of IMT during training on learning efficiency and later retention and the effect of IMT during recall appear not to have been clearly determined. Accordingly, the present study was designed as a further test of the effects of: (*a*) IMT on the efficiency of acquisition of verbal material, (*b*) IMT during training on the retention of verbal material, and (*c*) IMT during recall on the recall of verbal material.

Method. Sixty undergratuate students from an introductory course in psychology at the University of Sydney acted as *S*s. Fifteen *S*s were randomly assigned to each of four cells of a 2 × 2 factorial design. The four groups of *S*s underwent a serial verbal learning task, and recall trials 10 min. after learning the list (to a criterion of one errorless trial) and 1 wk. later. The groups differed according to whether the acquisition and recall trials took place under conditions of IMT or MR as follows: Group 1 (acquisition IMT–recall IMT); Group 2 (acquisition IMT–recall MR); Group 3 (acquisition MR–recall IMT); and Group 4 (acquisition MR–recall MR).

The *S*s were seated in a sound-reduced room separated from *E* and the apparatus. Stimuli were recorded on a Hanimex tape recorder and delivered to *S* via headphones. A hand dynamometer was attached to the left arm of *S*'s chair. Activity of the forearm flexor muscles was continuously recorded for each *S* via silver–silver chloride disc electrodes attached to the left forearm after preparation with conducting gel. Recordings were made on a Grass IIID polygraph set to record at a paper speed of 1.5 cm/sec and calibrated at 25 μv. = 1-cm. pen deflection. An automatic switching device rewound the tape after each presentation of the list and restarted it 60 sec. after the previous presentation had ceased. An ear plug attached to the output channel of the tape recorder permitted *E* to monitor stimulus presentations.

The stimuli consisted of the following 12 disyllabic words chosen on the basis of their low polarity values (Heisse, 1965): EVENT, STUDENT, OFFICE, EFFECT, EXPRESS, HUSBAND, RECORD, PRODUCT, SPIRIT, NOTICE, MORNING, RESULT. There was a 2-sec. interval between stimulus presentations.

Upon entering the laboratory, each *S* was instructed that the task entailed learning a serial list of 12 words. The *S* was then seated in an armchair in the sound-reduced room, and headphones and electromyogram (EMG) electrodes were fitted. Further instructions were relayed to *S* via an inter-

---

[1] Requests for reprints should be sent to Helen C. Beh, Department of Psychology, University of Sydney, Sydney, 2006, New South Wales, Australia.

TABLE 1

MEAN TRIALS AND ERRORS TO CRITERION AND MEAN ERRORS AT 10-MIN. AND 1-WK. RECALL

Group	Trials to criterion	Errors to criterion	Errors at 10-min. recall	Errors at 1-wk. recall
1 (IMT–IMT)	8.1	34.3	1.0	2.5
2 (IMT–MR)	8.0	34.6	1.7	4.2
3 (MR–IMT)	9.6	41.8	.9	5.0
4 (MR–MR)	9.7	40.0	.4	5.4

Note. Abbreviations: IMT = induced muscle tension; MR = muscular relaxation. First abbreviation gives condition during acquisition; second gives condition during recall.

com system connecting the cubicle and the laboratory. The *S*s were instructed they would hear a short tone followed by the list of words, after which another tone would be presented to signal that they were to recall the list in serial order during the 60-sec. interval between presentations. This procedure was repeated until *S* had reached a criterion of one errorless trial. Omissions and words out of order were scored as errors. In the conditions where *S*s were required to exert muscle tension, instructions were given to grip the dynamometer attached to the chair at half maximum strength. This level was monitored by polygraph recording throughout the session. The EMGs of the MR groups were also continuously monitored to ensure that these *S*s maintained MR during the acquisition trials. Feedback from *E* permitted *S*s to correct any deviations from the required degree of MT, although this was required in only two cases.

Following the acquisition trials, each *S* was presented with a form of the Ravens Progressive Matrices Test, which was administered to prevent rehearsal between acquisition and the recall test 10 min. later. After 9 min. had elapsed since the final acquisition trial, *S* was asked to stop doing the test, to relax or grip the dynamometer (according to grouping), and to orally recall the list previously learned. Only one recall trial was given, and *S*s were asked prior to leaving if they would return at a later time to participate in the same sort of experiment which required learning letter–number pairs rather than a serial list, and an appointment was made for a corresponding time 1 wk. later. The following week EMG electrodes were attached and the headphones fitted; *S* was then asked to recall (during IMT or MR according to grouping) the list learned the previous week in serial order. Only one trial was given and *S*s were questioned afterward concerning rehearsal of the list during the preceding week. The seven *S*s who admitted to rehearsal were not included in the study. For the group of *S*s used, trials and errors to criterion were scored for the acquisition phase of the study, and errors were scored for the 10-min. and 1-wk. recall sessions.

Results. Table 1 shows the mean trials and errors to criterion and the mean errors during recall trials over groups. The mean errors on recall trials are presented graphically in Figure 1.

Four analyses of variance, each incorporating three planned contrasts, were carried out on these data. Analysis of the trials-to-criterion data revealed that groups learning with IMT took significantly fewer trials to reach criterion than groups under MR during the acquisition session, F (1, 56) = 9.52, $p < .01$. Differences between Groups 1 and 2 (both learning under IMT) and Groups 3 and 4 (both learning under MR) were not significant. Analysis of errors to criterion and errors at 10-min. recall revealed no significant differences. For the errors at 1-wk. recall, however, the difference between IMT-trained groups was significant, F (1, 56) = 7.75, $p < .01$, while differences between Groups 1 and 2 and between Groups 3 and 4 were not significant.

Discussion. From the obtained results it may be seen that while IMT during acquisition results in more efficient learning in terms of trials taken to reach criterion, there is no significant difference between IMT and MR groups when errors to criterion are considered. The finding of no difference between groups at the 10-min. recall interval indicates that where a list has been learned to criterion, IMT during either training or testing trials has no significant effect on retention at this interval. Further, the present results indicate that IMT during recall does not facilitate long-term (1-wk.) recall. The significant difference between IMT- and MR-trained groups at the 1-wk. recall interval, however, suggests that IMT during acquisition does enhance long-term recall.

While the results of this study are essentially in agreement with those of Bills (1927, Experiment I), who reported that IMT during training resulted in fewer trials to criterion and better retention of a serial list of verbal material, any apparent conflict in findings between the present study and previous studies may possibly be explained in terms of procedural differences. For example, with the exception of the study by Bourne (1955), studies scoring errors after a given number of trials consistently report fewer errors for the IMT-trained groups. In the present study, however, errors to criterion were scored rather than errors after a given

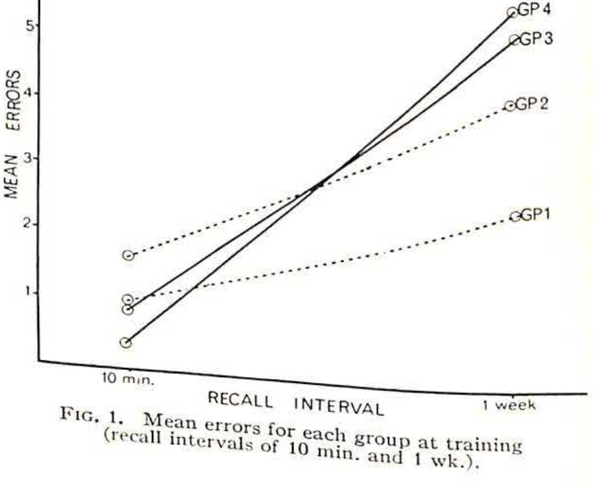


FIG. 1. Mean errors for each group at training (recall intervals of 10 min. and 1 wk.).

number of training trials, and thus the findings based on error scores in this study are not strictly comparable with findings based on error scores in other studies. The various studies of the effect of IMT on learning and retention have differed not only in the selected measure of the dependent variable, but also in terms of the type of verbal learning task employed (serial or paired associate), the number of items used, the difficulty of items used, the number of training trials given, and the time lapse between training and testing trials. It is suggested that such differences in procedure among the various studies may account for the conflicting findings, and that further studies need to be undertaken to examine the role of such factors in learning and retention under IMT.

REFERENCES

BILLS, A. G. The influence of muscular tension on the efficiency of mental work. *American Journal of Psychology*, 1927, **38**, 227–251.

BOURNE, L. E. An evaluation of the effect of induced tension on performance. *Journal of Experimental Psychology*, 1955, **49**, 418–422.

COURTS, F. A. Relations between experimentally induced muscular tension and memorization. *Journal of Experimental Psychology*, 1939, **25**, 235–256.

HEISSE, D. R. Semantic differential profiles for 1,000 most frequent English words. *Psychological Monographs*, 1965, **79** (8, Whole No. 601).

STAUFFACHER, J. C. The effect of induced muscular tension upon various phases of the learning process. *Journal of Experimental Psychology*, 1937, **21**, 26–46.

UEHLING, B. S., & SPRINKLE, R. Recall of a serial list as a function of arousal and retention interval. *Journal of Experimental Psychology*, 1968, **78**, 103–106.

(Received February 22, 1972)

(*Continued from page 205*)

Visual and Verbal Coding in Short-Term Memory. D. J. Murray (Department of Psychology, Queen's University, Kingston, Ontario, Canada) and Frances M. Newman.

Relationships among Short-Term Recall, Intralist Intrusions, Subjective Certainty Ratings, and Long-Term Memory. Anthony F. Grasha (IRTHE, University of Cinicinnati, 3020 Woodside Place, Cincinnati, Ohio 45221), Donald A. Schumsky, and Lee A. Elliott.

Memory for Physical and Semantic Features of Visual Material in a Shadowing Task. Ralph Hall (Department of General Studies, University of South Wales, P. O. Box 1, Kensington, New South Wales, Australia 2033), Diana Swane, and R. A. Jenkins.

An Information Processing Analysis of Visual Imagery. Douglas Griffith (Department of Psychology, University of Utah, Salt Lake City, Utah 84112) and William A. Johnston.

Resistance to Extinction at Spaced Trials Using the Within-Subject Procedure. Roger L. Mellgren (Department of Psychology, 455 West Lindsey, Room 705, University of Oklahoma, Norman, Oklahoma 73069) and Jeffrey A. Seybert.

Input Order and Output Interference in Organized Recall. Anderson D. Smith (School of Psychology, Georgia Institute of Technology, Atlanta, Georgia 30332).

Transfer of Incidental Learning to Free Recall. Robert E. Hicks, Mary T. Tarr, and Robert K. Young (Department of Psychology, Mezes Hall 211, University of Texas at Austin, Austin, Texas 78712)

Facilitation Effects of Incomplete Reward Reduction in Discrimination: A Within-Subject Evaluation. Richard L. Patten (Department of Psychology, Iowa State University, Ames, Iowa 50010).

Verbal Discrimination Learning and Two-Category Classification Learning as a Function of List Length and Pronunciation Instructions. John J. Shaughnessy (Department of Psychology, Northwestern University, Evanston, Illinois 60201).

A Comparison of Forward versus Backward Recognition Masking. Dominic W. Massaro (Department of Psychology, University of Wisconsin, Charter at West Johnson, Madison, Wisconsin 53706).

Failure of Subliminal Word Presentations to Generate Interference to Color Naming. Laurence J. Severance (Department of Psychology, Duke University, Durham, North Carolina 27706) and Frederick N. Dyer.

Stimulus Location as a Factor in Associative Symmetry. William H. Tedford, Jr. (Department of Psychology, Southern Methodist University, Dallas, Texas 75222) and J. Stephen Hazel.

Patterned versus Unpatterned Sequences of Study and Recall Trials in Free Recall of a Categorizable Word List. James G. Simmons (Department of Psychology, University of Rhode Island, Kingtson, Rhode Island 02881).

Effects of Amount of Interpolated Learning and Length of Retention Interval upon Retroactive Inhibition in a Serial Search Task. Robert E. Hicks and Robert K. Young (Department of Psychology, Mezes Hall 211, University of Texas at Austin, Austin, Texas 78712).

Short-Term Representation of Sentences. Robert Weisberg (Department of Psychology, Temple University, Philadelphia, Pennsylvania 19122).

Pronunciation and Apparent Frequency in a Between-Subject Design. Larry Wilder (Communication Arts, 1025 West Johnson Street, University of Wisconsin, Madison, Wisconsin 53706), Joel R. Levin, Elizabeth S. Ghatala, and Sandra McNabb.

(*Continued on page 222*)

Journal of Experimental Psychology
1973, Vol. 98, No. 1, 209–211

SELECTIVE ATTENTION IN THE EFFECT OF LABELING ON SYMBOLIC PROBLEM SOLVING

JOHN E. NORTHMAN[1] AND HOWARD RANKEN

Purdue University

Training with verbal labels encoding problem-relevant information facilitates subsequent solution of mental jigsaw problems in which stimuli can be fitted together to form specified configurations. Forty *S*s were assigned to four treatment conditions in a 2 × 2 design using two sets of stimulus shapes and two conditions of labeling. Labels encoding relevant figural information also facilitated performance on mental jigsaw problems only when the jigsaw pieces contained additional irrelevant information. The facilitating effect of labels is apparently due to selective attention, rather than to effects on memory or ease of symbolic manipulation.

Several studies (e.g., Ranken, 1963; Ranken & Dowling, 1965; Ranken & Wang, 1965; Saltz & Newman, 1960) attempted to define the role of labeling in problem solving by investigating conditions under which labels may help or hinder *S* in arriving at a solution. The results of these studies indicate that training with verbal labels, which enables *S* to encode problem-relevant information, markedly facilitates subsequent solution of jigsaw problems requiring *S* to decide how the stimulus shapes can be fitted together to form a specified configuration. Furthermore, this effect is known to be more pronounced for symbolic problems, where *S* must manipulate the stimulus shapes "in his head," than for concrete problems, where *S* has the shapes before him (Ranken & Dowling, 1965).

It is thus evident that under certain conditions labels facilitate symbolic problem solving. The question then arises as to why labels aid in problem solving. The following investigation evaluated the contribution of two possible sources of facilitation suggested by Ranken and Dowling (1965): (*a*) the effect of label training in inducing selective attention to problem-relevant information during training; and (*b*) the effect of labels on memory for figural information and on ease of symbolic manipulation of such information.

To separate these two classes of effects, two levels of labeling—a Named and an Unnamed condition—were combined with two sets of stimulus shapes—"two-sided" and "four-sided" shapes. These sets differed in the amount of irrelevant information which the shapes contained, and therefore, in the extent to which selective attention could operate. The two-sided shapes varied only in their left and right sides, each of which was either straight, pointed to the left, or pointed to the right. The top and bottom contours were always straight. The four-sided shapes had the same left and right contours as the two-sided shapes, but top and bottom contours, each of which was either straight, pointed up, or pointed down, also varied.

The labels encoded only the left and right sides for both sets of stimulus shapes. These were the only sides which had to be fitted together to solve the jigsaw problem. Since the preproblem training with the two-sided shapes required discrimination of these contours, and the shapes contained no other information, it was assumed that no opportunity for differential selective attention existed, and that any facilitation produced by labels would reflect their effects on memory or symbolic manipulation, but not on selective attention.

With four-sided shapes it was assumed that training with the labels, which encoded only the left and right sides, might induce selective attention to these problem-relevant contours, and that any facilitation produced by labels would reflect their effect on selective attention, in addition to the effects on memory and symbolic manipulation. Therefore, the difference between Named and Unnamed conditions with the two-sided shapes would be a measure of the effect of labels on memory and symbolic manipulation; while the increase in this difference as the number of sides increased from two to four—the interaction of naming with number of sides—would be a measure of the effect of labels on selective attention.

Method. The *S*s were 40 undergraduate psychology students randomly and equally assigned to the four treatment conditions: Named–two-sided, Unnamed–two-sided, Named–four-sided, Unnamed–four-sided.

Two sets of nine shapes each, shown in Figure 1, were used as stimuli. Each shape was based on a 1-in. square. Individual shapes were presented on 4-in. white square cards; the entire set of nine shapes also appeared on 7-in. white square cards.

In a preliminary presentation phase, the shapes were shown to *S* and their structure was explained. The *S*s in the Named condition were given two-digit labels which encoded each of the left and right sides of the shapes as follows: 1 = points to the left; 2 = straight; 3 = points to the right. (For example, in Shape 12 the left side points to the left and the right side is straight.) The *S*s in the Named condition were told to think of each shape in terms of its number; *S*s in the Unnamed condition were told to think of each shape simply as a shape. The nine shapes were then presented, one at a time, for 10 sec.

---

[1] Requests for reprints should be sent to John E. Northman, who is now at the Department of Psychology, State University of New York College at Cortland, Cortland, New York 13045.

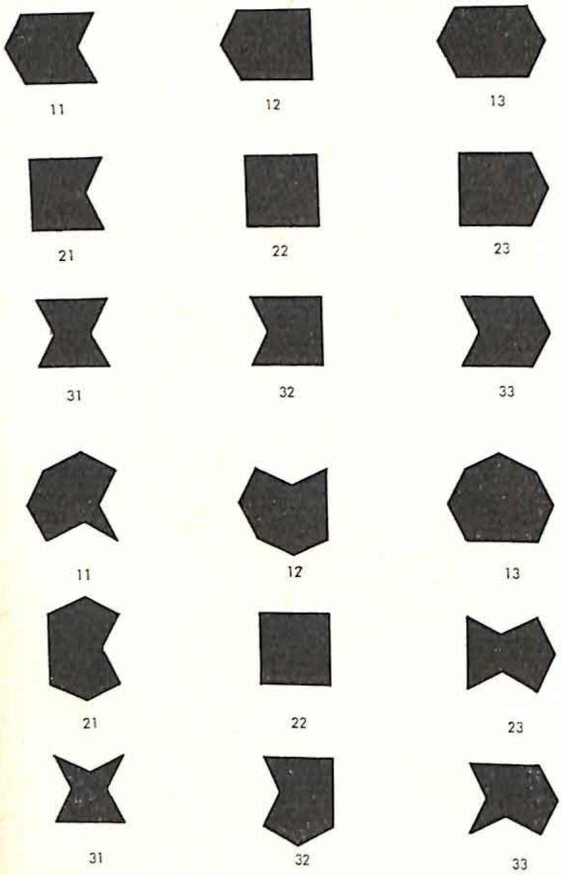


FIG. 1. Top: two-sided shapes; bottom: four-sided shapes.

each; the label for each shape was given for *S*s in the Named condition.

All *S*s then received training in a recognition task, designed to provide an opportunity in the Named condition to practice the labels, while giving *S*s in the Unnamed condition equal practice in discriminating the shapes without using words. A sample shape was shown for 2 sec.; after a 2-sec. pause, the entire set of nine shapes was presented, and *S* was asked to point to the shape that matched the sample. The *S*s in the Named condition were also asked to give the label, with correction, for each sample shape before making the recognition response. Training for all *S*s continued to a criterion of two consecutive errorless trials for all nine shapes on the nonverbal recognition response and, for the Named condition, also on the labeling response.

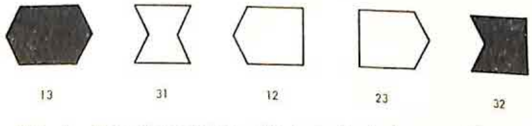


FIG. 2. Sample problem with two-sided shapes. (The *S* is presented with the shapes in the first and fifth positions; he must think of three other shapes [illustrated in figure] which could be fitted between these two, jigsaw fashion, to form a complete row.)

TABLE 1

MEAN NUMBER OF PROBLEMS CORRECTLY SOLVED BY EACH *S*

Condition	Figure		Both
	Two sided	Four sided	
Named			
M	12.6	12.4	12.5
SD	2.06	1.85	
Unnamed			
M	11.1	6.4	8.75
SD	2.34	2.58	
Both			
M	11.85	9.4	10.625

Each *S* was then given 18 jigsaw problems in which he was shown two shapes, which were to be used as the first and fifth shapes in a row of five. The *S*'s task was to think of three other shapes which could be fitted between these two, jigsaw fashion, to form a complete row. (See Figure 2 for an example.) It was stipulated that no shape, including the first and fifth shapes, could be used more than once in a given problem. The *S*s in the Named condition were told to think of the shapes in terms of their numbers; *S*s in the Unnamed condition were told to think of the shapes simply as shapes. When *S* indicated that he had solved the problem, the entire set of nine shapes was shown and *S* was tested by designating the shape that belonged in the middle position in the row. The problems were designed so that there was only one correct solution for the first 6 problems and two correct solutions for the remaining 12 problems. Timing for each problem began when *S* was presented with the first and fifth shapes and terminated when *S* said ready, indicating that he had arrived at a solution. Any *S* who did not solve a given problem within 180 sec. was given that time for a latency score and credited with an incorrect choice response.

Results and discussion. The mean number of problems solved correctly per *S* for each of the four experimental conditions is presented in Table 1. There was little variation in performance among *S*s in three of the four treatment conditions: Named–two-sided, Unnamed–two-sided, and Named–four-sided. The *S*s in these three groups performed noticeably better than *S*s in the Unnamed–four-sided condition. An analysis of variance showed a

TABLE 2

MEAN SOLUTION TIME (IN SEC.) OF PROBLEMS SOLVED CORRECTLY

Condition	Figure		Both
	Two sided	Four sided	
Named			
M	32.48	46.61	39.54
SD	17.00	14.16	
Unnamed			
M	27.86	50.60	39.23
SD	14.59	35.70	
Both			
M	30.17	48.60	39.39

significant Name × Side interaction, F (1, 36) = 9.20, $p < .01$. Analyses of simple effects indicated that only the simple effects involving the Unnamed–four-sided condition were significant. Because of the differential performance of *S*s in the Unnamed–four-sided condition, there were significant overall effects of both names and sides. The *S*s in the Named condition correctly solved significantly more problems than *S*s in the Unnamed condition, F (1, 36) = 25.55, $p < .001$, and *S*s working with two-sided shapes correctly solved significantly more problems than *S*s working with four-sided shapes, F (1, 36) = 10.91, $p < .01$.

The mean solution times of problems solved correctly for each experimental condition are shown in Table 2. An analysis of variance indicated that only sides yielded a significant effect, F (1, 36) = 6.19, $p < .05$, with *S*s working with two-sided shapes arriving at correct solutions significantly faster.

The effect of naming on the number of problems solved correctly was greater for *S*s working with four-sided than with two-sided shapes. It is suggested that this effect resulted, at least in part, from the names inducing selective attention to the contours subsequently relevant for solving the jigsaw problems. Performance in the Named–four-sided condition was as good as in the two-sided conditions; this indicates that selective attention, suggested by Ranken and Dowling (1965) as a possible effect of labeling, was sufficiently strong to eliminate any distracting effect of the irrelevant contours. It is possible, however, that there is a more complex relationship between procedural instructions and the effectiveness of labels. For example, the procedure used in this study for inducing selective attention appears to be an effective one; *S*s in the Named condition were given labels for the stimulus shapes and were encouraged to use them throughout the procedure. The effects of labels may be diminished with different instructions, e.g., in situations where labels are presented but their use is not so strongly encouraged.

The lack of any reliable effect of naming with two-sided shapes indicates that the names had a less facilitating effect on either memory for contours or ease of symbolic manipulation, the other possible effects of labeling suggested by Ranken and Dowling (1965). The time data further indicate that, in problems of this type, labels do not significantly accelerate finding a correct solution. An explanation for these results may lie in the nature of the two-sided shapes. It is conceivable that *S*s working with the simpler two-sided shapes found no need for assistance in memory from a device such as labels; a purely imaginal representation may have been sufficient. Indeed, several *S*s in the Named–two-sided condition indicated afterward that they found the labels to be of no assistance in solving the jigsaw problems. It appears, then, that problems with limited stimulus information can be handled at a perceptual level without any need for symbolic encoding. The importance of labels in problem solving increases with additional stimulus information, and this facilitating effect is at least partly due to labels inducing differential selective attention to relevant stimulus attributes during training.

REFERENCES

RANKEN, H. B. Language and thinking: Positive and negative effects of naming. *Science*, 1963, **141**, 48–50.

RANKEN, H. B., & DOWLING, W. J. Language and thinking The interaction of naming with relevance and concreteness. *Psychonomic Science*, 1965, **3**, 459–460.

RANKEN, H. B. & WANG, H. S. Language and thinking: The interaction of amount of pretraining and relevance in the effect of naming on problem solving. *Psychonomic Science* 1965, **3**, 461–462.

SALTZ, E., & NEWMAN, S. E. The effect of prior learning of symbols on performance in reasoning. *American Journal of Psychology*, 1960, **73**, 91–99.

(Received March 2, 1972)

Journal of Experimental Psychology
1973, Vol. 98, No. 1, 212–213

EFFECT OF SEQUENCE STRUCTURE ON RECALL[1]

BARRY J. SCHWARTZ,[2] DANIEL S. LORDAHL,[3] AND BLASE GAMBINO

University of Miami

Recall for sequences containing 8 and 10 binary events was assessed in 100 *S*s. Sequences were presented in either a serial or simultaneous mode. Errors were analyzed in terms of the structure of the sequences, including the presence of runs and alternations, the number of transitions, and the code length for the sequence. Code lengths of the beginning and ending parts of the sequences were studied along with the code for the entire sequence. Group coding differences are discussed.

In two recent reviews of the binary-prediction research (Jones, 1971; Myers, 1970), the structure of the event sequence and *S*'s memory for prior events have been amply documented as powerful variables in the determination of *S*'s predictions. In addition, current explanations of binary prediction have emphasized such processes as encoding or chunking (cf. Myers, 1970) of the event sequence. Presumably, *S*'s ability to recall the past sequence of events should be related to the ease with which the sequence can be encoded or chunked. The present study was designed to explore this relationship in a task which did not involve prediction, but involved only recall.

Method. The *S*s were 100 students from introductory psychology classes at the University of Miami, participating for experimental credit. Experimental sessions involved small groups of between three and six persons each, and lasted for approximately $\frac{1}{2}$ hr.

The sequences were binary event sequences with 1 and 2 being the events. Twenty-five Length 8 and 25 Length 10 sequences were used.

There were two experimental conditions: serial and simultaneous, with 50 *S*s in each. The *S*s in both conditions were tested first with the 25 Length 8 sequences, followed by the 25 Length 10 sequences. The sequences for both groups were identical; only the mode of presentation differed. Five different orderings of the sequences were used, each order being used for 10 *S*s in each condition. Assignment of the groups of *S*s to a condition and sequence order was random, with the only limitation being the filling of the groups.

The 50 *S*s in the serial condition had the sequences presented to them one digit at a time by means of a sound recording. The sequences were read at the rate of 1 digit/sec. The tape was stopped between sequences in order to give *S*s enough time to write down their answers. When all of the members of a group had indicated that they had finished writing, the next sequence was presented. It is important to point out that the taped sequences were read with as little emphasis and intonation as possible. Digits were carefully timed for presentation at precisely every second, with emphasis being put on the last digit only.

The 50 *S*s in the simultaneous condition had the sequences presented to them by an opaque projector so that they saw all of the digits simultaneously. The sequences were projected onto a screen from unlined 3 × 5 cards which had been carefully prepared so that they could be seen clearly; the room was lighted so as to permit projection of the sequences, but still let *S* see well enough to write down his answers. Each sequence was exposed for a time equal to 1 sec/digit. As with the serial group, the next sequence was not presented until all *S*s had indicated that they were ready.

The instructions were identical for both conditions, except for the specific details concerning the mode of presentation of the sequences. The *S*s were told that the purpose of the experiment was to create a new scale relating intelligence to memory for digits. They were told that they would see or hear a sequence of the numbers 1 and 2, and that it was their task to write down as much of the sequence as they could remember as soon as presentation of the sequence had been completed. The fact that there was no penalty for guessing was emphasized. Answer sheets with spaces for the 8- and 10-digit sequences were provided to make sure that *S*s responded with the correct number of digits. The sequences were then presented with no warm-up or practice trials.

Results and discussion. Defining an error as a sequence which was not completely correct, the two groups were compared in a repeated-measures analysis of variance with the two sequence lengths as the repeated measure. Recall was more difficult under serial presentation than under simultaneous presentation ($p < .001$). Since *S*s in the simultaneous group were able to read the entire sequence in 3–5 sec., their superiority may have been due to having an additional 5–7 sec. to rehearse and/or encode the sequence. This analysis also revealed that as sequence length increased, recall decreased ($p < .001$); and this deterioration was enhanced under the serial mode ($p < .001$). This finding of increased difficulty in recall with longer sequences was, of course, expected. Another factor which may have entered into these results pertains to the number of transitions within each sequence.

---

[1] Portions of this article were presented at the 1969 meeting of the Psychonomic Society in St. Louis. This study was submitted by the first author in partial fulfillment of the requirements for the master's degree at the University of Miami.

[2] Requests for reprints should be sent to Barry J. Schwartz, Department of Psychology, University of Miami, P. O. Box 8185, Coral Gables, Florida 33124.

[3] Now at the Florida State University.

A transition is defined as the number of times which two successive digits within a sequence differ. In grouping the sequences according to the number of transitions, it was found that the number of errors was highest on those sequences with a moderate number of shifts (i.e., four transitions). Sequences with fewer shifts tended to be more accurately recalled, possibly as a result of their having long runs. On the other hand, those sequences with a large number of transitions are also easier to remember because they resemble single alternations. These results indicate that as run lengths become shorter, recall diminishes; but at a certain point the sequence begins to approximate a single alternation pattern and the process is reversed.

In another repeated-measures analysis of variance, this time using each sequence as one of the repeated measures, the sequence term was consistently found to be significant at the .001 level. This analysis was done separately for each combination of sequence length and presentation mode. This finding supports the conclusion that errors accrued to certain sequences due to the structural properties of those sequences. On the basis of this conclusion, a more detailed analysis of structure was conducted by classifying all sequences according to the patterns formed by the first and second half of each sequence.

In all conditions, sequences beginning with a run, single alternation, or double alternation had fewer mean errors per sequence. A point-biserial correlation of mean errors with whether or not the first four digits formed a run or alternation was significant only for the serial Length 8 group ($p < .01$). No similar relationship was found to exist between errors and ending patterns. This finding suggests that the beginning patterns may be more critical to recall, perhaps functioning as anchor points for the remainder of the sequence.

If *S*s are encoding sequences, then recall may be related to the length of the code which they use, sequences with shorter codes being recalled more readily. Using Morganstein's (1970) definition of code length, errors were correlated with code length for each combination of sequence length and mode of presentation.[4] A Spearman correlation was used, and the code lengths represented either the entire sequence or the coding of the beginning or the ending of the sequence. Table 1 presents these correlations and their significance levels.

TABLE 1

SPEARMAN CORRELATIONS BETWEEN SEQUENCE CODE LENGTHS AND NUMBER OF ERRORS

Sequence length	Mode of presentation	Code length		
		Entire	Beginning	Ending
8	Serial	.36*	.89**	.01
	Simultaneous	.45*	.41	.00
10	Serial	.50**	.82*	.32
	Simultaneous	.50**	.18	.78*

* $p < .05$
** $p < .01$.

From Table 1 it can be seen that the code length of the entire sequence is significantly correlated with recall errors under all four conditions. When code length is derived solely from the beginning pattern, the correlation improves for the serial *S*s under both sequence lengths. A similar increase in the size of the correlation is found for the simultaneous Length 10 condition for the ending code length.

The finding of significant correlations of errors with the code length of the sequence as a whole supports Morganstein's (1970) earlier work. In addition, the presence of larger correlations for the beginning codes under the serial mode fortifies the point-biserial correlation previously discussed and points toward the preeminence of the beginning pattern for this group. The increased size of the correlation based on the ending pattern for the simultaneous Length 10 condition is more difficult to explain, especially since no such relationship was found to exist for the Length 8 sequences. It may be that the addition of two digits strains the coding and/or memory processes, favoring the recall of those sequences containing easily coded ending patterns. The code lengths of the beginning patterns has no relationship with recall for this group, presumably due to the fact that the first portion of the sequence can be more thoroughly rehearsed or more readily stored independent of its code length. The left-to-right scanning by *S*s undoubtedly plays a role here.

In summary, the data suggest that recall is related to sequence structure. Possible mediating factors include the presence of runs and alternations and the length of a code used in the recall process. Differences between serial and simultaneous modes point to the coding of different portions of the sequences under the two conditions, with emphasis on the coding of the beginning portion for serial *S*s and on the ending portion for *S*s in the simultaneous condition.

REFERENCES

JONES, M. R. From probability learning to sequential processing: A critical review. *Psychological Bulletin*, 1971, **76**, 153–185.

MORGANSTEIN, S. Effect of encoding on short-term memory. *Journal of Experimental Psychology*, 1970, **86**, 387–392.

MYERS, J. L. Sequential choice behavior. In G. H. Bower & J. T. Spence (Eds.), *The psychology of learning and motivation: Advances in research and theory*. Vol. 4. New York: Academic Press, 1970.

(Received April 12, 1972)

---

[4] The authors wish to express their appreciation to the Consulting Editor for his helpful suggestions with regard to this analysis.

ournal of Experimental Psychology
973, Vol. 98, No. 1, 214–216

HOMOGRAPHS AS DOUBLE FUNCTION ITEMS IN VERBAL DISCRIMINATION LEARNING[1]

DONALD H. KAUSLER[2]

University of Missouri at Columbia

Partial double-function list conditions were constructed for verbal discrimination learning with homographs serving as the double-functioning words (i.e., as a right item in one pair and as a wrong item in another pair). Pairing the homographs with associates related to their distinctive meanings resulted in faster acquisition than pairing the homographs either with associates related to a common meaning or with words unrelated to the homographs. The results were interpreted in terms of a feature-tagging hypothesis.

According to the feature-tagging hypothesis of verbal discrimination learning (Kausler, in press), intrapair discrimination between right (R) and wrong (W) items requires a comparison between the stored representation of the R item plus its frequency-of-response tag (f_R) and the stored representation of the W item plus its frequency-of-response tag (f_W). To the extent that $f_R > f_W$, *S* selects the R item, in agreement with the basic tenets of Ekstrand, Wallace, and Underwood's (1966) frequency theory. Considerable research on other facets of recognition learning has indicated that the stored representation of a word consists, in turn, of a set of semantic and sensory features, with each feature being tagged separately by a frequency cue. False recognitions occur when the tagged features of "old" study words overlap with either the semantic or the sensory features of "new" test words. A comparable process seemingly operates in verbal discrimination learning when R items share either semantic or sensory features with interpair W items (e.g., Fulkerson & Kausler, 1969). That is, *S* falsely recognizes W words as R words when features of the W words have been tagged with f_R cues elsewhere in the list content, thereby increasing the difficulty of learning the overall list.

The foregoing analysis accounts readily for the extreme difficulty of learning a completely double-function list, that is, one in which each R item also serves as a W item for another pair in the list (Kausler & Boka, 1968). In effect, the usual consequence of double-functioning items is to create a situation in which f_R and f_W become indistinguishable as cues for intrapair discrimination, and *S* must adapt to the situation by relying upon other sources of cuing if he is to learn the list (Kausler & Boka, 1968). The difficulty stems from the complete overlap of features between R and W items that occurs when ordinary words are double functioning. However, the foregoing analysis also suggests that double-functioning words qua words need not have double-functioning sets of features. That is, the features of a word that are tagged in its function as an R item may be segregated, at least in part, from the features that are tagged in its function as a W item. If this segregation is accomplished, a double-function list should be reduced to the functional equivalence of a single-function list.

Homographs, or words with two or more clearly discriminable meanings (e.g., Meaning A and Meaning B), represent an ideal word population for testing the segregation hypothesis. Recent research has revealed that correct recognition of a homograph as an "old" word is markedly reduced when the study context implies Meaning A and the test context implies Meaning B. Similarly, the false recognition of a "new" test word as "old" is greater when that word is related to Meaning A (the study-context meaning) than when it is related to Meaning B (cf. Perfetti & Goodman, 1970). When applied to the double-function verbal discrimination situation, an R item related to Meaning A should be discriminable from a W item related to Meaning B, even though the items have completely overlapping sensory features. The present design called for a modified double-function list in which only half of the words, viz., the homographic content, served both R and W items. The remaining words functioned singly as either R or W items.

In the primary experimental condition, each homograph was paired with an associate linked to Meaning A of that homograph when it functioned as an R item (e.g., NET–TENNIS as a R–W pair) and with an associate linked to Meaning B when it functioned as a W item (e.g., NET–INCOME as a W–R pair). It was anticipated that the differential associative contexts would trigger the different meanings for tagging purposes, thus reducing the list to single-functioning items. In another experimental condition, each homograph was paired with an associate of a single meaning of that homograph, both when it functioned as an R item and when it functioned as a W item. Two different groups served under this condition. In the one group the single meaning was the primary, or more frequently responded to, meaning (Kausler & Kollasch, 1970) of each homograph (e.g., NET–TENNIS as a R–W pair and NET–CATCH as a W–R pair). In the other group, the single meaning was the secondary, or less frequently responded to, meaning of each homograph (e.g., NET–GAIN as an R–W and NET–INCOME as a W–R pair). In both cases it was anticipated that the homographic items would retain their double functioning as R and W items, thus making the lists difficult to learn. In the control condition each homograph was paired with an unrelated word, both when it functioned as an R item (e.g., NET–REST as an R–W pair) and when it functioned as a W item (e.g., NET–CHEAT). It seems likely that an *S* who perceives a homograph in terms of its primary (or secondary) meaning in one pairing will perceive it similarly in the second pairing. Thus, double functioning of homographs was the likely consequence in the control condition. However, there was the possibility that idiosyncratic relationships with the paired words would result in differential encodings of some homographic items across pairs, thus reducing the overall double-functioning nature of the list.

---

[1] Thanks are due Greg Settle for his assistance in various phases of this study.

[2] Requests for reprints should be sent to Donald H. Kausler, Department of Psychology, University of Missouri, Columbia, Missouri 65201.

TABLE 1

Summary Statistics for Total Errors and for Errors across Blocks of Trials

Group	Trial block						Total	
	1–5		6–10		11–15			
	M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
M	20.05	7.35	8.80	6.87	5.00	7.20	33.85	19.20
P	21.10	6.84	16.00	8.38	10.80	7.32	47.90	20.40
S	24.00	7.20	15.05	8.92	11.45	11.08	50.50	24.80
C	20.95	5.96	13.60	7.69	8.60	7.53	43.15	18.14

Note. Abbreviations: M = mixed meanings; P = primary meaning; S = secondary meaning; and C = control.

Method. The *S*s were 80 general psychology students; all were naive with respect to prior participation in verbal learning experiments, and all received extra course credit for participation in the experiment. They were assigned randomly within blocks of four *S*s to the following groups: Group M (mixed meanings of the homographs in R and W functions); Group P (the primary meaning of the homographs in both R and W functions); Group S (the secondary meaning of the homographs in both R and W functions); and Group C (control condition, without cuing of either meaning).

For purposes of greater generalization of results and greater control over idiosyncratic characteristics of homographic items, two forms (I and II) of each list condition were constructed. Two sets of six homographs each, selected from the Kausler and Kollasch (1970) word association norms, formed the core items for constructing specific lists. The homographs in Forms I and II had primary meanings averaging 66.6% and 81.9%, respectively, of the total associations to the homographic stimuli. The two forms were subsequently found to yield neither a main effect on error rate nor an interaction effect with list conditions. Consequently, the two forms were pooled for the analyses reported later. For each condition (12 R–W pairs) the six homographs were R items in 6 pairs and W items in the remaining 6 pairs. The intrapair items yoked to each homograph (i.e., W items for R-functioning homographs and R items for W-functioning homographs) consisted of low-strength (less than 10% in the Kausler & Kollasch norms) associates to that homograph in Groups M, P, and S, and words of zero-strength association in Group C. For Group M the two associates paired with each homograph were scored in the Kausler and Kollasch norms as referring to the separate meanings of that homograph, whereas for Group P and Group S the two associates were scored in the norms as referring to the same meaning (primary and secondary, respectively). For Group C, Form I condition, the two words paired with each homograph were the two words paired with one of the homographs in the Form II condition of Group M; for Group C, Form II, the selections came from the Form I condition of Group M. In addition to associative relatedness to the homographic items, the following criteria guided the specific selection of nonhomographic items: (*a*) avoidance of direct associative relatedness with one another, as determined by existing word association norms, when applicable, and by *E*'s judgment otherwise; (*b*) balancing the cross-list conditions with respect to frequency and grammatical form class of words serving as items. Especially important was the avoidance of direct associative relatedness between the two nonhomographic words paired with a given homograph. Selection of the nonhomographic words without regard to their associative relatedness could have resulted in differential extents of double functioning across conditions through implicit associative responses. Word associates linked to a common meaning are more likely to be directly related associatively than associates to the distinctive meanings. Since implicit associative responses to these related words as R items would add frequency units to their W item counterparts (Ekstrand et al., 1966), the overall difficulty of the lists for Groups P and S would be greater than that of the lists for Groups M and C, independently of the segregation of meanings problem. Thus, the nonhomographic items paired with a common meaning were carefully selected so as to be both associates of the homograph and independent of one another. For example, one of the homographic items was NET. In Group P the words paired with NET were TENNIS and CATCH, relatively weak associates of NET that are seemingly associatively unrelated to one another.

The procedure followed closely that of Kausler and Boka (1968) in terms of temporal conditions (anticipation method, 2:2 sec. rate, 4-sec. intertrial interval), left–right positioning of items (with R items underlined during feedback intervals), and controls for serial ordering effects and spatial position effects. The initial trial was a guessing trial, and it was followed by 15 learning trials.

Results and discussion. Means and standard deviations for total errors over the 15 learning trials are given in Table 1 for all four groups. Errors were also analyzed separately for Trials 1–5, Trials 6–10, and Trials 11–15. Means and standard deviations for the trial blocks are included in Table 1. The error scores (square-root transformation; pronounced skewed distributions) were subjected to a 4×3 mixed analysis of variance, with homographic conditions as the between-group variable and trial blocks as the within-group variable. Both the main effect for homographic conditions and the main effect for trials were significant, $F\ (3,\ 76) = 2.78$, $p < .05$; and $F\ (2,\ 152) = 166.62$, $p < .001$, respectively. An orthogonal components analysis revealed a significant comparison in total errors between Group M and the other three groups combined, $F\ (1,\ 76) = 4.63$, $p < .05$. Neither the Groups P and S vs. Group C nor the Group P vs. Group S comparison approached significance. In addition, individual contrasts between means (Duncan's test) yielded significant differences between Group M and Group P ($p < .05$) and between Group M and Group S ($p < .05$). The contrasts between Group M and Group C and between Group C and Group S approached significance ($ps < .10$). The remaining contrasts were clearly nonsignificant ($ps < .10$). Further evidence for the overall superiority of Group M came from the fact that 12 *S*s in this group attained a criterion of two consecutive perfect trials, as compared to four in Group P, seven in Group S, and seven in Group C.

The Conditions × Blocks interaction effect was also significant, $F\ (6,\ 152) = 3.30$, $p < .005$. From Table 1 it may be seen that the interaction resulted from differential learning rates between Group M and the remaining groups. Separate analyses of variance yielded a significant between-group effect for Trials 6–10, $F\ (3,\ 76) = 5.34$, $p < .01$; and for Trials 11–15, $F\ (3,\ 76) = 4.31$, $p < .01$; but not for Trials 1–5 ($F < 1$). In addition, orthogonal comparisons for the last two blocks of trials revealed that only the comparisons between Group M and the other three groups combined were significant, $F\ (1,\ 76) = 14.83$, $p < .001$; and $F\ (1,\ 76) = 11.75$, $p < .001$.

The present results offer support for the feature-tagging hypothesis. The difficulty of learning a list containing double-functioning words can be markedly reduced when separate features of these words discriminate between their use as R and W items in a verbal discrimination list. However, there is reason to believe that the list condition for Group M was not equivalent to that of a single-function list condition in which completely different words serve as R and W items. That is, the learning rate for Group M was considerably slower than the rate ordinarily found for single-function groups receiving 12-pair lists. For example, Fulkerson and Kausler (1969) found that all 20 of their *S*s in the control group attained a criterion of two consecutive perfect trials in less than 15 trails, with the mean number of trials being 7.60. It will be recalled that only 12 of the 20 *S*s in the present study attained this criterion, and moreover, the mean number of trials for these *S*s was 10.75. The failure to eliminate double functioning completely is not surprising in that the orthographic and phonemic features of the homographs overlapped completely in their double functions, even though the semantic features were presumably discriminable. As noted in the introduction, false recognition of W words as R words increases when sensory features of the W words have been tagged with f_R cues elsewhere in the list. The present results also offer little support for the hypothesis that *S*s spontaneously segregate homographs into their distinctive meanings during practice on double-function pairs. While Group C displayed fewer errors than Groups P and S, the advantage was slight and did not reach statistical significance.

REFERENCES

Ekstrand, B. R., Wallace, W. P., & Underwood, B. J. A frequency theory of verbal-discrimination learning. *Psychological Review*, 1966, **73**, 566–578.

Fulkerson, F. E., & Kausler, D. H. Effects of intrapair and interpair bidirectional associates on verbal-discrimination learning. *Journal of Verbal Learning and Verbal Behavior*, 1969, **8**, 307–310.

Kausler, D. H. Orthographic distinctiveness of homonyms and the feature tagging phyothesis of verbal discrimination learning. *American Journal of Psychology*, in press.

Kausler, D. H., & Boka, J. A. Effects of double functioning on verbal discrimination learning. *Journal of Experimental Psychology*, 1968, **76**, 558–567.

Kausler, D. H., & Kollasch, S. F. Word associations to homographs. *Journal of Verbal Learning and Verbal Behavior*, 1970, **9**, 444–449.

Perfetti, C. A., & Goodman, D. Semantic constraint on the decoding of ambiguous words. *Journal of Experimental Psychology*, 1970, **86**, 420–427.

(Received May 1, 1972)

Journal of Experimental Psychology
1973, Vol. 98, No. 1, 217–219

EFFECTS OF INSTRUCTION AND STIMULUS PRESENTATION ON THE OCCURRENCE OF AVERAGING RESPONSES IN IMPRESSION FORMATION[1]

HARRY F. GOLLOB[2]
University of Denver

ANDREW M. LUGG
University of Michigan

In 1965 Anderson found that adding moderately polarized adjectives to highly polarized adjectives decreased the polarity of *S*s' responses. To test whether the occurrence of such averaging responses was enhanced by Anderson's methodology, a replication experiment was performed. One factor of a 2 × 2 factorial design, which followed Anderson in requiring responses at fixed intervals, compared Anderson's instructions with instructions which were similar but did not specify that the adjectives were equally important. The second factor compared Anderson's method of presenting adjectives in isolation with a sentence form of presentation. A fifth condition allowed *S*s to proceed at their own rate through the stimuli. Anderson's basic results were replicated. No significant effects due to instructions, format of presentation, or timed vs. untimed presentation were obtained.

Anderson (1965) obtained evaluative ratings of compound stimuli constructed from various combinations of moderately polarized and highly polarized adjectives of the same evaluative sign. He reported that his *S*s gave *averaging* responses, that is, their evaluation of a compound stimulus was *less* extreme than their evaluation of the most extreme component stimulus. An *adding* response was scored whenever an *S*'s evaluation of a compound stimulus was *more* extreme than his evaluation of the most extreme component stimulus.

The study reported in this article is a replication of Anderson's (1965) study with variations in the instructional set, the format of stimulus presentation, and the amount of time allowed for rating each set of adjectives. We expected these variables to affect the strength of adding or averaging responses given by *S*s. Using Anderson's adjective sets as stimuli, we hypothesized that, semantic and syntactic considerations aside, some of our conditions would lead *S*s to give averaging responses and other of our conditions would lead *S*s to give adding responses.

In an attempt to avoid stimulus interaction, differential weighting, and discounting effects, Anderson used instructions which included the paragraph: "It is important that you pay *equal attention* to each adjective. Occasionally the adjectives may seem inconsistent, but you should consider each adjective as *equally important* in describing the person [our italics]."[3] Moreover, just before *S*s began the experimental booklet they were reminded to "pay equal attention to each adjective." Although the use of such instructions may help ensure the validity of the scaling techniques used by Anderson, we hypothesized that such instructions lead *S*s to average their subjective impressions. In particular, we suggest that emphasizing equal importance of the adjectives is similar in effect to telling *S*s that each of the individual traits best characterizes the stimulus person. Since it is not possible that more than one trait simultaneously best characterizes a stimulus person, we felt that Anderson's instructions enhanced the probability of *S*s inferring that some conflict existed between the pieces of information, and that *S*s resolved the conflict by taking the average of their evaluative impressions.

Hendrick's (1967) results provide some support for the above line of reasoning. He did three experiments and obtained the strongest averaging results in the only experiment which instructed *S*s to "consider the one or two traits listed as the most characteristic traits the person possesses [p. 64]." This experiment also differed from the other two in that *S*s were required to rate one item every 15 sec. The three studies also differed in other less important ways.

Averaging also may have been enhanced by the format of stimulus presentation used by Anderson. His stimuli were presented in the following manner: Each page of a test booklet contained two adjectives on each of one or two lines, together with a blank space in which *S*s recorded their ratings of how much they liked or disliked the person described by the set of adjectives. Partly on the basis of pilot results using two-adjective sets, we speculated that if the adjectives were presented as components in sentences (e.g., "The person is nonconforming, warmhearted, deliberate, and considerate."), *S*s would tend to average less than they do when faced with isolated adjectives. We felt that *S*s would find the sentence presentation more "natural," and that they would therefore assume that the description of the stimulus person simply presented a number of traits each of which captured some aspect of the person's character but none of which best characterized him. In such contexts, we hypothesized, *S* does not treat the adjectives as equally important pieces of information to be weighed

---

[1] This study was supported by National Institute of Mental Health Grant 18177. We thank Norman H. Anderson for sending us the detailed materials which were necessary in order to replicate his 1965 study and for making some helpful comments concerning the present paper. We also thank Betty B. Rossman for her useful criticisms of an earlier draft of the paper.

[2] Requests for reprints should be sent to Harry F. Gollob, Department of Psychology, University of Denver, Denver, Colorado 80210.

[3] This quotation was taken directly from the instructions used by Anderson in his 1965 study.

against each other, but rather attempts to build up an overall picture of the stimulus person. If the new information is *confirming*, *S* will tend to like the stimulus person more, whereas if the new information is *disconfirming*, he will tend to like the stimulus person less. A similar statement applies in the case of negatively evaluated stimuli.

In order to partially test some of the above ideas a 2 × 2 factorial experiment was conducted. One of the factors compared Anderson's instructions vs. instructions which did not specify that each adjective was equally important nor that equal attention should be paid to each adjective. The other factor compared presentation of the adjectives in isolation vs. the sentence form of presentation. We expected to obtain the most averaging in the cell which used Anderson's instructions and method of stimulus presentation, and to obtain adding (or at least less averaging) in the cell which used neither Anderson's instructions nor his method of stimulus presentation.

In Anderson's (1965) study, *S*s were required to spend a total of 15 sec. on each item before giving it a rating. We felt that this length of time was exceptionally long and that this may have affected the results. Thus, in addition to the four groups described above, we ran a condition in which *S*s were allowed to proceed at their own rate through the booklet, were presented with stimuli in sentence form, and were given instructions which did not tell them to pay equal attention to each adjective nor that each adjective was equally important. We considered this condition the most natural and expected it to yield more adding than any of the four timed conditions.

Method. The stimuli and method of constructing test booklets were identical to those used by Anderson (1965). The component adjectives were selected from among those which Anderson (on the basis of normative data) classified as highly liked (H), moderately liked ($\mathrm{M^+}$), moderately disliked ($\mathrm{M^-}$), and highly disliked (L) personality-trait adjectives. Each of the four stimulus replications included four different sets each of HH, $\mathrm{M^+M^+}$, $\mathrm{M^-M^-}$, and LL adjectives. Within each replication these pairs were combined with each other in a way (described by Anderson) that yielded two different sets each of the types HHHH, $\mathrm{M^+M^+M^+M^+}$, $\mathrm{M^-M^-M^-M^-}$, and LLLL; four different $\mathrm{HHM^+M^+}$ sets, and four different $\mathrm{LLM^-M^-}$ sets. Thus, a single stimulus replication contained a total of 32 sets of adjectives, half being two-adjective sets and half being four-adjective sets.

TABLE 1

Analysis of Variance of Adding–Averaging Scores

Source	*df*	*F*
Mean	1	6.88*
Condition (C)	4	1.39
Replication (R)	3	1.60
CR	12	1.29
S(CR) (Mean Square)	60	(59.92)
Evaluation (E)	1	<1.00
CE	4	1.24
RE	3	<1.00
CRE	12	1.50
S(CR) × E (Mean Square)	60	(50.59)

* $p < .05$.

Each *S* received a booklet of 10 practice items and a booklet of 32 experimental items. Each page of the booklets contained a single item and a blank space for *S* to record his response. The pages of each booklet were arranged in a different scrambled order for each *S*.

The format of presentation was either the isolated adjective format used by Anderson or a sentence format. In the isolated adjective format, each page of a test booklet contained two adjectives on each of one or two lines; whereas in the sentence format each page of a booklet contained a sentence with either two or four adjectives. For example, "The person is dependable, unconventional, friendly, and meditative."

The *S*s in the condition that was a straightforward replication of Anderson's study were read the same instructions which Anderson used. The instructions used in our other conditions closely paralleled those used by Anderson. In our "not equally important" conditions, we simply deleted the portions of Anderson's instructions which emphasized *S*s should give equal importance and equal attention to each of the adjectives. Depending on whether the isolated adjective format (i.e., that used by Anderson) or the sentence format of presentation was used, and on whether the condition was timed or untimed, very minor changes in wording were effected in order to make the instructions appropriate for each experimental condition. We quote from Anderson (1965) to describe his basic procedure:

> The *S* was instructed to read the adjectives slowly to himself. A total of 15 sec. was used for each set. At the end of the allotted time, *E* said, "Rate & Turn." The *S* then recorded his response and turned to the next page.
>
> The *S*s were told to use 50 to rate a person they would neither like nor dislike, to use lower numbers for persons they would dislike, and to use higher numbers for persons they would like [p. 396].

The *S*s were 80 male students who participated in the study in order to fulfill research participation requirements in their introductory psychology course. The *S*s were run one, two, or three at a time and were randomly assigned in equal numbers to each of the 20 groups (i.e., 4 Stimulus Replications × 5 Conditions).

Results. The mean rating given to all of the HH, $\mathrm{HHM^+M^+}$, $\mathrm{LLM^-M^-}$, and LL items was computed for each *S*. Two "adding–averaging" scores were computed for each *S* by subtracting the *S*'s HH mean from his $\mathrm{HHM^+M^+}$ mean, and his $\mathrm{LLM^-M^-}$ mean from his LL mean. These scores have the property that averaging responses yield negative scores and adding responses yield positive scores. The mean adding–averaging scores for the conditions using Anderson's instructions with isolated adjectives vs. sentences, respectively, were

−1.46 and −3.93; the means using "not equally important" instructions with isolated adjectives vs. sentences, respectively, were −1.50 and −1.75, and the mean for the untimed condition which used sentence presentation was .62. The adding–averaging scores were analyzed by use of a three-way analysis of variance in which the five experimental conditions (C) and the four stimulus replications (R) were between-*S* factors, the evaluation (E) of the adjectives being a repeated measurement factor. As shown in Table 1, none of the main effects or interactions were significant. In order to look more closely at the data, we pretended that we had made a correct a priori prediction as to which condition would yield the greatest and which the smallest adding–averaging score. The differences between the means for the conditions were so small relative to their variability, that even this "planned" comparison (see Winer, 1971, pp. 478–484) did not yield a significant F value. Moreover, the difference between the weak adding result obtained in the untimed condition and the mean of the averaging results obtained in the four timed conditions was not significant. The grand mean of −1.60, however, was statistically significant ($p < .02$) in indicating that averaging results were obtained.

We also replicated Anderson's (1965) finding of a highly significant set-size effect, i.e., an increase in the polarity of the evaluative rating given to a compound as the number of nearly equally evaluated components in the compound stimulus increases. Thus, the mean of the adding–averaging scores, which were computed for each *S* by taking the sum of the differences, (HHHH) − (HH) and the analogous differences for the M^+, M^-, and L adjectives, was 5.61, $F\ (1, 60) = 95.84$, $p < .0001$. The size of the effect was not significantly affected by either conditions or replications, or by the evaluation of the adjectives.

In sum, we found an averaging effect and a set-size effect as did Anderson (1965). We did not, however, find any significant effects due to instructions, format of presentation, or timed vs. untimed presentation.

Discussion. It is apparent that our specific expectations concerning the effects of instructions and format of stimulus presentation were not met. Perhaps more powerful manipulations of these factors would have revealed an effect. It also may be that averaging effects dominate when large sets of adjectives are considered, and that the factors discussed above make a difference when smaller sets of two or three adjectives are considered. Be that as it may, the averaging result seems to be a more robust phenomenon than we had thought.

REFERENCES

Anderson, N. H. Averaging versus adding as a stimulus-combination rule in impression formation. *Journal of Experimental Psychology*, 1965, **70**, 394–400.

Hendrick, C. A. *Averaging versus summation in impression formation.* (Doctoral dissertation, University of Missouri) Ann Arbor, Mich.: University Microfilms, 1967, No. 68-303.

Winer, B. J. *Statistical principles in experimental design.* (2nd ed.) New York: McGraw-Hill, 1971.

(Received May 6, 1972)

Journal of Experimental Psychology
1973, Vol. 98, No. 1, 220–222

EFFECTS OF ASSOCIATIVE REACTION TIME AND MEANINGFULNESS IN FREE RECALL OF MIXED AND UNMIXED LISTS[1]

RONALD LEY[2] AND JEFFREY DEAN

State University of New York at Albany

Two groups of 12 *S*s each were tested for free recall on four lists of eight CVCVCs each. One group was tested on four mixed lists: lists balanced with respect to short- and long-latency (SL and LL) associative reaction time (RT) units within levels of high and low meaningfulness (HM and LM). The other group was tested on four unmixed lists; each list was homogeneous with respect to RT and M, and the same 32 CVCVCs were used as for the mixed-list group. Recall performance of the two groups was not different, and the list factor did not interact with either RT or M. High-M units were recalled more frequently than low-M units, but the effect of RT was limited to low-M units.

Recent experiments on the effects of associative reaction time (RT) in paired-associate learning and free recall (e.g., Ley, 1968; Ley & Locascio, 1970; Locascio & Ley, 1972a) have employed mixed lists, i.e., lists composed of short- and long-latency RT words and paralogs within levels of high and low meaningfulness (M). This mixed-list procedure has been criticized on the basis of what Underwood (1966) has called priority habits; Underwood has proposed the hypothesis that *S*s, through past experience, may develop performance habits which differentially affect their learning of different types of verbal units. Thus, the results of experiments dealing with the effects of stimulus characteristics on learning may be a function of priority habits (functional stimulus differences) rather than of the characteristics of the stimuli *E* manipulates (nominal stimulus differences) if a mixed list (heterogeneous verbal units within a list) is used rather than an unmixed list (homogeneous verbal units within a list).

Although there has been little systematic study on the issue of the effects of mixed vs. unmixed lists, it is tenable that the consistent recall superiority of short-latency RT units over long-latency RT units, reported by Ley and his associates (e.g., Ley, 1968; Ley & Locascio, 1970; Locascio & Ley, 1972a), may be a function of the mixed list.

The purpose of the present study was to determine whether the mixed list (lists containing short- and long-latency RT units within two levels of M) is a factor in free recall. If the mixed list is a factor in free recall, then free recall of units presented in unmixed lists (lists containing the same verbal units used in the mixed lists, but homogeneous with respect to RT and M for a given list) should differ from free recall of units presented in a mixed list. Furthermore, if the mixed list is in part responsible for the effects of RT within levels of M, then an interaction between type of list and RT might be expected.

Method. The *S*s were 24 graduate students from the State University of New York at Albany (13 males and 11 females) enrolled in a graduate course in educational psychology. All *S*s were naive with respect to previous experience in verbal learning studies.

The materials were 32 CVCVC verbal units selected from the 319 units assessed for scaled meaningfulness (m') by Locascio and Ley (1972b) and for RT by Taylor and Kimble (1967). Using m' as the primary criterion, 16 CVCVCs were selected which had short-latency (SL) RT values, 8 from the high-meaningfulness (HM) end of the m' distribution and 8 from the low-meaningfulness (LM) end of the m' distribution; and 16 CVCVCs were selected which had long-latency (LL) RT values, 8 from the HM end and 8 from the LM end of the m' distribution.

All verbal units with an m' value of 2.00 or less were chosen as representative of LM. The RT values of this LM sample were then taken from Taylor and Kimble's (1967) list. The mean ($\bar{X}$) and standard deviation (*SD*) of these RT values were calculated. Verbal units which fell one or more *SD*s below the $\bar{X}$ were considered for the purposes of the present study to be SL–LM units, and verbal units which fell one or more *SD*s above the $\bar{X}$ were considered to be LL–LM units. Eight CVCVC verbal units were then selected from each of these RT groups. The same procedure was used in selecting the HM–SL and HM–LL verbal units. The m' and RT values for each of the 32 CVCVCs selected are given in Table 1, from which it can be seen that the mean m' values of the SL and LL RT verbal units are essentially the same within both HM and LM levels.

The 32 CVCVCs were divided into four mixed lists, balanced with respect to SL and LL RT within levels of HM and LM; and four unmixed lists, homogeneous with respect to RT and M. The mixed lists were constructed by randomly dividing the 32 CVCVCs into four lists, with the restriction that each list contain 2 CVCVCs from each of the four classifications (HM–SL, HM–LL, LM–SL, and LM–LL). The order of the CVCVCs on the four mixed lists was random with the qualifications that (*a*) each classification group had one unit in each

---

[1] This research was supported in part by a State University of New York Research Foundation grant-in-aid to the first author. The authors wish to thank Edward Tesiny for his assistance in the collection and analysis of the data.

[2] Requests for reprints should be sent to Ronald Ley, Department of Psychology, State University of New York at Albany, Albany, New York 12222.

TABLE 1

Mean Associative Reaction Time (RT), Mean Rated Associative Frequency (a'), and Mean Scaled Meaningfulness (m') for the 32 CVCVC Verbal Units of the Mixed and Unmixed Lists

RT	CVCVC	RT	a'	m'
High meaningfulness Short latency	TENOR	1.88	4.05	3.86
	DIGIT	1.80	4.00	3.80
	RAVEN	1.78	4.00	3.80
	TOTEM	1.84	3.58	3.26
	DOZEN	1.81	4.26	4.16
	PATER	1.90	3.20	2.87
	LILAC	1.91	4.10	3.92
	DIVAN	1.86	2.94	2.55
Long latency	FAVOR	2.80	3.94	3.70
	BASIS	2.52	3.72	3.42
	HABIT	2.70	3.94	3.70
	RADAR	2.55	4.08	3.90
	RUMOR	2.80	4.23	4.12
	LAGER	2.68	3.54	3.23
	VICAR	2.82	3.26	2.92
	CYNIC	3.35	3.56	3.25
Low meaningfulness Short latency	JUREL	2.39	1.72	1.09
	GORAL	2.62	1.64	.97
	RATAL	2.64	1.66	1.00
	LIMEN	2.40	1.69	1.05
	NEROL	2.77	1.41	.52
	SALOL	2.70	1.54	.82
	TONUS	2.29	1.69	1.05
	BOLUS	2.78	1.66	1.00
Long latency	ZUREN	4.65	1.48	.70
	XABIN	3.86	1.87	1.27
	MENAD	3.72	1.64	.97
	VUTAW	4.26	1.36	.36
	HYRAX	4.04	1.66	1.00
	NIDUS	3.71	1.62	.94
	JEHAD	3.85	1.66	1.00
	POLEF	3.74	1.84	1.23

TABLE 2

Mean Number of CVCVCs of Short- and Long-Latency Associative Reaction Time (RT) and Low and High Meaningfulness Recalled Under Mixed and Unmixed List Conditions

RT	List			
	Mixed		Unmixed	
	M	*SD*	*M*	*SD*
Low meaningfulness Short latency	2.75	1.42	2.58	1.31
Long latency	2.17	1.70	1.67	1.07
High meaningfulness Short latency	5.08	1.56	4.92	1.73
Long latency	5.00	1.35	5.00	1.41

half of the list, (*b*) no two units of the same classification group appeared consecutively on a list, and (*c*) the position occupied by units in each classification group in each half of the list was different for the four lists. The unmixed lists were constructed by dividing the same 32 CVCVCs into four lists corresponding to the four classifications (HM–SL, HM–LL, LM–SL, and LM–LL). The order of the eight CVCVCs, homogeneous with respect to RT and M, was random within each unmixed list.

A practice list, containing eight different CVCVCs, was constructed in the same manner as the four mixed lists. The practice list was similar to the mixed lists, i.e., the units were selected from the remaining CVCVCs assessed for m' and RT. All CVCVCs were typed in capital letters and photographed for presentation on a Kodak Carousel slide projector.

The *S*s were instructed to try to learn the 32 CVCVCs in a free-recall task under a silent rehearsal mode of study.

An equal number of *S*s ($n = 12$) was randomly assigned to each of the two list conditions: mixed lists and unmixed lists. The *S*s were tested in groups of three by the same *E*. The three *S*s of each group were seated in front of a screen and read standard instructions explaining the study procedure to be followed and the recall procedure which would be employed. The mixed practice list was then presented to *S*s with each unit presented at a 4-sec. rate by means of a timed slide presentation.

Following the presentation of the practice list and study by *S*s under the instructed procedure, the projector was stopped and *S*s were given 90 sec. to print the presented units of the list on the appropriate page of a $5\frac{1}{2} \times 8\frac{1}{2}$ in. booklet. Immediately following the recall period for the practice list, the four test lists were presented. Each test list was presented at a 4-sec. rate, followed immediately by a 90-sec. recall period. A 20-sec. interval was employed between the end of one recall period and the presentation of the next test list. The *S*s were instructed to print the CVCVCs recalled on the pages of a booklet provided by *E*. The order of the four lists was counterbalanced so that each list was presented first, second, third, or fourth an equal number of times. This procedure also provided for a counterbalancing of sequence. If a unit was correctly printed by *S*, a score of one was given; if a unit was incorrectly printed, a score of zero was given.

The design employed was a mixed factorial in which list conditions (mixed or unmixed) was a between-*S*s factor, and RT (SL, LL) and m' (HM, LM) were within-*S*s factors.

Results. The mean numbers of CVCVCs recalled under the mixed and unmixed list conditions are given in Table 2. The differences between the mean number of units recalled under the mixed list was not significantly different from the mean number of units recalled under the unmixed list, F (1, 22) $= .27$, $p > .05$. Furthermore, the list factor did not interact with RT, F (1, 22) $= .04$, $p > .05$; with M, F (1, 22) $= .30$, $p > .05$; or with the M × RT interaction, F (1, 22) $= .18$, $p > .05$.

Consistent with the results of Ley and his associates, high-M units were recalled more frequently than low-M units, F (1, 22) $= 143.12$, $p < .001$. Although the overall effect of RT was not significant, F (1, 22) $= 2.66$, $p > .05$, short-latency RT units ($\bar{X} = 2.66$) were recalled more frequently than long-latency RT units ($\bar{X} = 1.97$), t (23) $= 3.95$,

$p < .001$. The RT × M interaction was not significant, $F(1, 22) = 1.59$, $p > .05$.

Discussion. The results of the present experiment do not provide any evidence that free recall of verbal units presented in mixed lists differs from free recall of the same units presented in unmixed lists. The absence of any interaction between the type of list and RT or M, and the remarkable similarity between recall scores of the mixed and unmixed lists, provide support for the validity of the mixed-list procedure used by Ley and his associates (Ley, 1968; Ley & Locascio, 1970; Locascio & Ley, 1972a) in their demonstrations of the effect of RT in verbal learning.

REFERENCES

Ley, R. Associative reaction time, meaningfulness, and presentation rate in paired-associate learning. *Journal of Experimental Psychology*, 1968, **78**, 285–291.

Ley, R., & Locascio, D. Associative reaction time and meaningfulness of CVCVC response terms in paired-associate learning. *Journal of Experimental Psychology*, 1970, **83**, 445–450.

Locascio, D., & Ley, R. Associative reaction time, meaningfulness, and mode of study in free recall. *Journal of Experimental Psychology*, 1972, **95**, 460–462. (a)

Locascio, D., & Ley, R. Scaled-rated meaningfulness of 319 words and paralogs previously assessed for associative reaction time. *Journal of Verbal Learning and Verbal Behavior*, 1972, **11**, 243–250. (b)

Taylor, J., & Kimble, G. Association value of 320 words and paralogs. *Journal of Verbal Learning and Verbal Behavior*, 1967, **6**, 744–752.

Underwood, B. J. *Experimental psychology*. (2nd ed.) New York: Appleton-Century-Crofts, 1966.

(Received May 18, 1972)

(*Continued from page 208*)

Effects of Pattern Goodness on Recognition Time in a Memory Search Task. Stephen F. Checkosky (Department of Psychology, Lake Forest College, Lake Forest, Illinois 60045) and Dean Whitlock.

Proactive Inhibition in the Short-Term Retention of Pictures. John C. Yuille (Department of Psychology, University of British Columbia, Vancouver 8, British Columbia, Canada) and Charles Fox.

Requiring Hypotheses and the Identification of Unidimensional, Conjunctive, and Disjunctive Concepts. Roger L. Dominowski (Department of Psychology, King's College, University of Aberdeen, Old Aberdeen AB9 2UB, Scotland).

Brief Storage of Compressed Digits. Dennis H. Holding, Emerson Foulke (Graduate School-Perceptual Alternatives Laboratory, University of Louisville, Louisville, Kentucky 40208), and Robert L. Heise.

Reproduction of Horizontal and Vertical Lines in a Within-Subject Design. Coleman T. Merryman (Department of Psychology, Mezes Hall 211, University of Texas at Austin, Austin, Texas 78712) and Sandra S. Merryman.

Effects of Amount of Training on the Type of Solution and the Breadth of Learning in Optional Shifts. Bryan E. Shepp (Department of Psychology, Brown University, Providence, Rhode Island 02912) and Marilyn J. Adams.

Auditory and Autonomic Tests of the Preparatory-Adaptive-Response Interpretation of Classical Aversive Conditioning. John J. Furedy (Department of Psychology, University of Toronto, Toronto 181, Ontario, Canada).

The "Uncertain" Response in Detection-Oriented Psychophysics. Charles S. Watson (Central Institute for the Deaf, 818 South Euclid, St. Louis, Missouri 63110), Steven C. Kellogg, David T. Kawanishi, and Patrick A. Lucas.

Stimulus Set, Response Set, and Word Identification. Graham R. Skanes (Institute for Research in Human Abilities, Memorial University, St. John's, Newfoundland, Canada) and Don C. Donderi.

Probability, Compatibility, Speed, and Accuracy. O. Joseph Harm (Department of Psychology, Vanderbilt University, Nashville, Tennessee 37240) and Joseph S. Lappin.

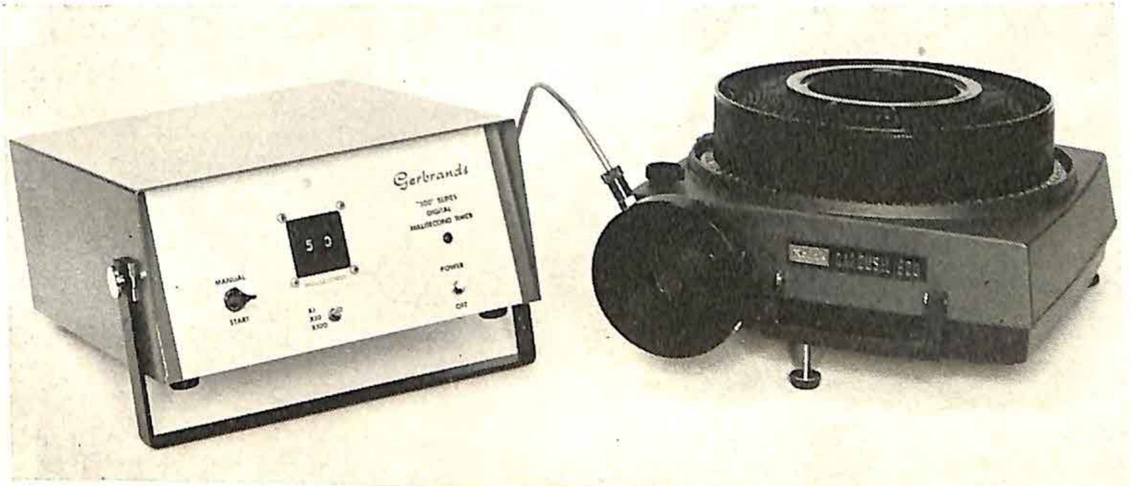
Nonlearning: The Completeness of the Blindness. Paul Fingerman and Marvin Levine (Department of Psychology, State University of New York at Stony Brook, Stony Brook, New York 11790).

Selection of Hypotheses as Affected by Their Preference Values. Diane F. Halpern and Francis W. Irwin (Department of Psychology, University of Pennsylvania, 3813–015 Walnut Street, Philadelphia, Pennsylvania 19104).

Space–Time Adaptation of Visual Position Constancy. John C. Hay (Department of Psychology, University of Wisconsin—Milwaukee, 3401 Downer Avenue, Milwaukee, Wisconsin 53201) and W. Mack Goldsmith.

Projection Tachistoscopes

featuring our new Electronic Tachistoscopic Shutters



- Available in 1 and 2 channel configurations with Digital Millisecond Timers and Kodak Carousel Projectors.
- Shutter with drive unit also available as a separate item. Can be used with any 35 mm slide projector and laboratory timer.
- Minimum exposure duration: 5 milliseconds.
- We can also provide a Slide Presentation Programmer with a variable interstimulus presentation interval. This unit automatically advances the Carousel (changing the slide) and presents a 100 millisecond warning tone 2 seconds before the stimulus presentation.

Write or phone for complete details.

Ralph Gerbrands Company
8 BECK ROAD, ARLINGTON, MASSACHUSETTS 02174 (617) 648-6415

"the standard of excellence in laboratories throughout the world"

Accept this FREE
144-page guide

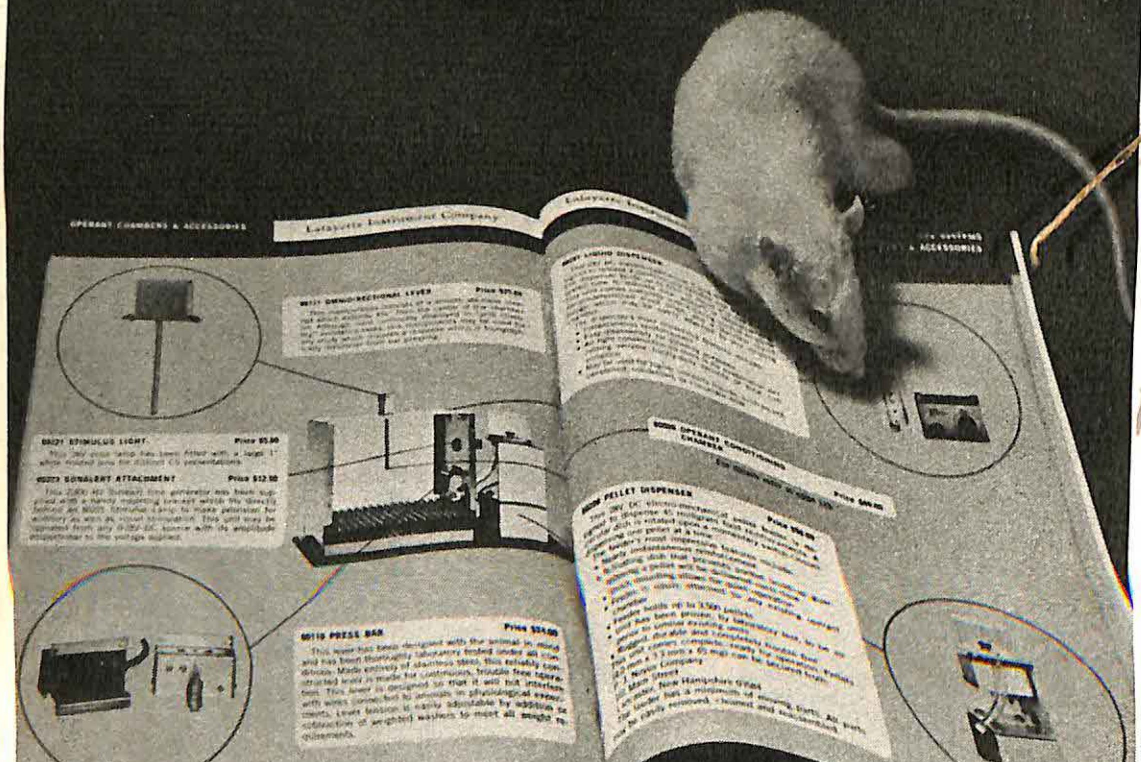
The new 1973 Lafayette Instrument Company guidebook of scientific and laboratory instruments for the life sciences.

On page 106, for instance, you'll see how to set up a custom student operant conditioning chamber. And that's just one page.

Each page is filled with facts, prices, and information about our newest quality instruments. So naturally, we'd like you to get your hands on our guidebook.

Send for your personal copy today. Ask for the 1973 edition of guidebook no. 1170.

LAFAYETTE INSTRUMENT COMPANY
P.O. Box 1279/Lafayette, Indiana 47902/ 317-423-1505



Library No. 718(3)
Date 321
Lib
Bureau Edn. Psy. Research

Journal of
Experimental Psychology

Copyright © 1973 by the American Psychological Association, Inc.

VOL. 98, NO. 2 MAY 1973

ROLE OF COGNITIVE FACTORS ON ADAPTATION TO PRISMATIC DISPLACEMENT[1]

JOHN J. UHLARIK[2]
Kansas State University

The *S*s were exposed to sensorimotor rearrangement consisting of prismatic displacement of the visual field. Independent exposure conditions differed in the kinds and amount of information regarding rearrangement available to *S*s. In addition, explicit instructions regarding the effects of looking through laterally displacing prisms manipulated *S*s' awareness of the rearrangement. It was found that qualitatively different kinds of information served equivalent functions, viz., shifts in localizations that were compensatory for prism-induced errors. However, the nature and magnitude of these adaptive shifts varied as a function of the source and amount of relevant information. The results were consistent with an information-processing formulation of the process of adaptation.

It has been well documented that *S*s can compensate for errors of localization induced by wedge prisms that laterally displace the visual field. However, the nature of the process underlying this adaptation has been subject to debate.

An information-processing formulation, suggested by a number of investigators (e.g., Freedman, 1968; Howard, 1970; Uhlarik & Canon, 1971; Wohlwill, 1966), explains the process of adaptation by the underlying assumption that adaptation develops as the result of accumulated information regarding the nature of the discordance produced by the sensory rearrangement. During conditions of norma sensorimotor coordination, inputs regarding spatial referents provide complementary or correlated information. Rearrangement represents instances of artificially induced conflict between typically correlated inputs such as vision and proprioception. Adaptation refers to the process by which the conflict is reduced or eliminated so that the inputs are once again congruent and the observer is capable of accurate sensorimotor coordination.

According to Wohlwill (1966), adaptation involves many inputs or channels, and what is critical is "the system of interrelationships linking those particular channels, rather than any particular channel such as vision or proprioception [p. 205]." In other words, information of any sort regarding the altered state of the system is hypothesized to be *sufficient* to serve as the basis for adaptation. In contrast, other positions focus on the *necessity* of a single kind of informational input or channel. For example, Held and Hein

---

[1] This research was supported by funds from Biomedical Sciences Support Grant FR-07036 from the National Institutes of Health and Grants F-665 and F-734 from the Bureau of General Research, Kansas State University. The author would like to thank Alan Osgood and Larry Pound for their assistance in conducting the experiment and Sam Brown and Robert Welch for their comments and criticisms on the manuscript.

[2] Requests for reprints should be sent to John J. Uhlarik, Department of Psychology, Kansas State University, Manhattan, Kansas 66506.

(1958) consider reafferent stimulation to be the single condition necessary for genuine adaptation. It is probable that the feedback provided by reafference is an excellent source of information sufficient to produce adaptation. However, Wohlwill (1966) suggests that "in its absence, other types, e.g., kinesthetic, or even purely cognitive, may serve an equivalent function [p. 205]." The broad purpose of the present study was to determine the functional equivalence of different kinds of information for the adaptive process.

The differences between various theoretical positions become more apparent upon examining the implications of a specific information-processing formulation. According to Canon (1970, 1971), the nature or locus of adaptation should vary as a function of which informational input is attended to during conditions of rearrangement. Inputs not attended to will be brought to be congruent with those to which *S*s are attending or using as the basis for making sensorimotor localizations. In the present study this information-processing formulation was tested by varying the different kinds of exposure conditions that *S*s experienced. The assumption was that the differences in training procedures represented qualitative and quantitative differences in the information available about rearrangement. One condition (cognitive feedback) provided information to *S* in the form of verbal feedback regarding errors of localization. The *S*s in this condition were required to make pointing localizations of visually displaced targets. However, these *S*s received no direct feedback regarding the discordance between the seen and felt positions of their limbs because they were never permitted to view their limbs while localizing targets during exposure. Rather, *E* merely informed *S*s of any errors of localization. Since these *S*s were unable to view their pointing during or after localization, the location of a target could not be specified in terms of visual afferent (reafferent) inputs arriving from that limb. Consequently, it was assumed that there was no intermodality discordance between the visual and proprioceptive inputs from the limb in this condition. It is important to note that by precluding *S*s' visual impressions of their limbs, the reafferent stimulation that Held (Held & Hein, 1958) would consider to be necessary for adaptation was eliminated.

A second exposure condition (concurrent display) primarily contained information regarding intermodality discordance. In this condition, *S* simply moved his limb back and forth on a pivot board. No targets were present during this period. Hence, there was no explicit information regarding errors of localization. This training procedure is referred to as a "concurrent display without targets" (Howard, 1968) because (*a*) *S* simultaneously sees and feels the position of his limb during exposure to rearrangement, and (*b*) there are no targets present to localize. The intermodality discordance in this condition stems from the discrepancy between the visual and proprioceptive inputs regarding the location of the limb. Held (e.g., 1968) has used almost exclusively this kind of exposure condition to provide support for his reafferent model of adaptation. As a result, the formulations proposed by Held and his associates (Hardt, Held, & Steinbach, 1971; Held, 1968) deal only with conditions of rearrangement that preclude direct error information regarding localizations. It should be possible to compare this condition, which should be optimal for producing adaptation according to Held's reafferent model, with the cognitive feedback condition, which should not produce adaptation if Held's theory of reafference is correct.

A third condition (terminal display) contained information regarding both error feedback and intermodality discordance. Here, *S* was required to localize visual targets and was able to view his pointing limb at the terminus of its extension toward the target. Thus, errors of pointing were apparent to *S* when he viewed his limb in relationship to the target. This training procedure corresponds to a "terminal display with target" (Howard, 1968).

It was predicted that all of the conditions would produce adaptive shifts in measures of pointing localizations. Furthermore, it was predicted that the magnitude of these shifts would vary as a function of the amount of information available to *S*s during exposure to rearrangement. Specifically, the adaptive shifts should be of greater magnitude for the terminal display condition than for either the cognitive feedback or concurrent display condition. This prediction was based on the assumption that the terminal display provided the greatest amount of available information in that *S*s received both target-error information *and* information regarding intermodality discordance, whereas in the cognitive feedback condition *S*s received only target-error information, and in the concurrent display condition *S*s received only information regarding intermodality discordance.

In studies of adaptation to prismatic displacement, *S*s may or may not be consciously aware of the presence of rearrangement. For instance, in the case of concurrent displays with no targets *S*s are typically unaware that the visual input has been transformed. Apparently the dominance of the visual input in this situation precludes explicit knowledge that there is a discordance between visual and proprioceptive inputs (Hay, Pick, & Ikeda, 1965; Rock, 1966). Conversely, the presence of target-error information under the terminal display and cognitive feedback conditions makes explicit information available to *S* that some aspect of normal sensorimotor coordination has been transformed. In order to control for these differences between conditions and to further investigate the importance of cognitive factors in prism adaptation, a second variable was manipulated: *S*s' awareness of the presence of prismatic displacement. Specifically, three levels of prism awareness were employed. At one level of prism awareness *S*s were informed of the nature of the rearrangement both during its introduction and its removal (complete prism awareness). In a second condition *S*s were informed of the rearrangement only during its removal (partial prism awareness). Under the third condition *S*s were not informed of the rearrangement either during its introduction nor during its removal (no prism awareness).

METHOD

Apparatus. The appratus is described in detail elsewhere (Uhlarik, 1972). Basically, it consisted of a horizontal panel mounted above the top surface of a table. The *S* was seated facing the apparatus, and pointing localizations were made by reaching under the panel. The panel had an opening that permitted *S* to view his limb during conditions that allowed visual feedback; the opening could be covered for conditions that preculded visual feedback.

The target display consisted of a vertical panel perpendicular to *S*'s line of sight. The panel was approximately 56 cm. from the frontal surface of the cornea of *S*'s right eye and was curved to a radius of 57 cm. The panel could be transilluminated with a movable target light that subtended a visual angle of .5°. The light could be transported on a track that was curved to the arc of the stimulus display panel and was mounted on rubber foam in order to preclude extraneous auditory cues during movement. The horizontal plane through which the light moved was 1.5 cm. above the top surface of the horizontal panel.

A second set of stationary targets consisted of five vertical lines that *E* could place on the back edge of the opening in the horizontal panel. These lines were 2 mm. wide, 5 cm. apart on center, and labeled A, B, C, D, and E. The lines extended down to the top of a coil that served as a position transducer.

The *S*s made localizations by attempting to touch that part of the position transducer that corresponded to the designated target. The position transducer was mounted on the side of the vertical stimulus display panel facing *S*; it was located 3 cm. below the horizontal panel and circumscribed the 100° arc of the stimulus display. The *S* wore a plastic finger splint taped to his right index finger. Pointing localizations were made by touching a metallic conductor on the end of the finger splint to the position transducer. This act completed an electrical circuit that provided a signal that was linearly proportional to the position of the stylus on the coil. The signal registered on a digital display voltmeter and was recorded by *E*, who was seated out of *S*'s field of view.

The *S* viewed the stimulus display with his head held in position by a biteboard throughout the experiment. When making the dental impression on the biteboard, *S* was instructed to position himself so that the center line, labeled C, appeared to be straight ahead of his right eye when viewed monocularly.

Visual displacement to the left or right for the right eye was produced by a Risley rotating prism (Bausch and Lomb No. 7148–59). During conditions of optical rearrangement the prism was adjusted to a strength of 20 diopters, which laterally displaced the visual field 11.3°. The prism was mounted in the right eyepiece of a pair of welder's goggles. The left eyepiece was occluded.

Experimental design and subjects. A 3 × 3 × 2 between-*S* factorial design was employed. The independent variables included three levels of prism awareness (complete prism awareness, partial prism awareness, and no prism awareness), three exposure conditions (terminal display, concurrent display, and cognitive feedback), and prism orientation (base left and base right). The latter variable was included to counterbalance direction of prismatic displacement. Three different dependent measures of egocentric localization (visuomotor, proprioceptive-motor, and visuo-proprioceptive-motor) were obtained.

One hundred and four college undergraduates served as *S*s in the experiment. Participation in psychological research permitted them to earn extra credit in the psychology course in which they were enrolled. Right-hand dominance and an absence of a history of astigmatism were prerequisites for the experiment. Four *S*s were randomly assigned to each cell of the factorial design. In addition, 16 *S*s were assigned to each of two control conditions to be described below. A total of eight *S*s were eliminated from the study and replaced by additional *S*s; one had difficulty with the biteboard and the other seven displayed ranges greater than 8° on pre- and/or postmeasures of visual or proprioceptive "straight ahead." This criterion for consistency of localizations was established prior to collection of the data.

Procedure. To familiarize *S* with the apparatus and the manner in which he was to make localizations, practice trials were given with the prism set at 0 diopters. The room lights were on and the opaque panel was removed so *S* could view his arm. The *S* made 15 pointing localizations by alternately touching two visible targets on the coil with the stylus attached to the right index finger. Next, *S* was given practice pointing straight ahead. The *S* was instructed to use his median or midsaggital plane as the referent for all straight-ahead localizations, i.e., to indicate what point appeared to him to be directly in line with the midline of his head and body or, in other words, straight ahead of the tip of his nose. Fifteen practice localizations were made in which *S* pointed straight ahead with eyes closed. At the end of each localization, *S* was allowed to open his eyes to determine how accurate he was. No scores were recorded during these practice localizations.

Next, a series of preexposure measures of three different types was made with the prism still set at 0 diopters. All pointing localizations were made with the right arm. Contact with the position transducer was made by reaching under the shoulder-level black panel that completely covered the horizontal surface of the apparatus. The *S*'s head was fixed in the straight-ahead position by the biteboard and no visual feedback from any part of the body was possible. The room was darkened except for a very dim light on *E*'s control panel shielded from *S*'s field of view.

Three different types of measurements were taken. One, termed visuomotor (VM), involved having *S* indicate the position he saw as straight ahead. The movable target light was illuminated in the periphery of *S*'s field of view, and *S* indicated the direction to move the light by pressing a left or right signal button. The *E* would move the light in the direction specified by *S*. There was no time limit on responding and *S* was instructed to repeat this procedure, including as many reversals in direction as were necessary, until the light appeared to be in the position that was straight ahead. This measure was repeated four times starting at alternate sides in the periphery of the monocular field.

Another measure, termed proprioceptive-motor (PM), involved the limb position that *S* selected as feeling straight ahead. It was obtained by having *S* point straight ahead with no visual target and no visual feedback from the pointing limb. The *S* was instructed to close his eyes and point straight ahead by touching the appropriate place on the coil with the stylus. The *E* recorded the reading on the meter and told *S* to return his hand to the surface of the table. This measure also was repeated four times.

The final measure required *S* to point at a series of visible targets, again with no visual feedback from the pointing limb. Since this measure would be expected to reflect both visual and proprioceptive effects, it was labeled the visuo-proprioceptive-motor index (VPM). The *E* illuminated the target light in the appropriate position, and *S* made his localization of it by touching the appropriate place on the coil. Then *S* signaled when he felt he was on target and then returned his hand to the table surface. Five target positions—0°, 6° left, 6° right, 12° left, and 12° right—were employed. The targets were presented in a randomly determined order. The target was presented once at each position with the exception of the 0° position which was presented twice.

The order of presentation of the VM and PM measures was counterbalanced. For half of the *S*s the VM measures were always taken first, and for the remaining half the PM measures were taken first. The VPM measures were always obtained last.

Following the preexposure measures *E* illuminated two 15-w. lights, located under the top surface of the apparatus, and adjusted the prisms to a strength of 20 diopters. For half of the *S*s displacement was to the left, and for the remaining half it was to the right. Three different conditions were used to inform *S*s of the effects of the optical rearrangement. In the no prism awareness condition *S*s were told to close their eyes while *E* made some adjustments in the equipment. The *S*s were told that the reason for keeping their eyes closed was to keep their state of dark adaptation constant. During this period *E* changed the strength of the prism from 0 to 20 diopters. The *E* made this adjustment while *S* was wearing the goggles; *E* viewed *S*'s right eye in a mirror to insure compliance with the instructions. Immediately before and after this adjustment *S*'s field of view contained only the homogeneous black surface of the unlit stimulus display panel, thus precluding the possibilities of *S*'s noticing a change in his visual environment after he opened his eyes. Following adjustment of the prism the apparatus was arranged so that the appropriate exposure targets were visible on the display, and the experiment continued. The same procedure was used for this condition following exposure to rearrangement when *E* adjusted the prism from 20 to 0 diopters.

A second condition, designated complete prism awareness, was designed to inform *S*s of the displacing effects of the prisms. In this condition *S* was told that the effect of the adjustment of the goggles was to displace the visual field to one side with a prism. Furthermore, *S* was instructed to keep his eyes open and watch as the field of target lines (which *E* placed on the display before the adjustment) moved to the left or right. The counterrotation of the two wedge prisms in the goggles produced a continuous displacement as the Risley prism was adjusted to 20 diopters. The same procedure was used for this condition following exposure to rearrangement when *E* adjusted the prism from 20 to 0 diopters.

A third condition, designated partial prism awareness, was designed to inform *S*s of the lateral displacement only at the end of exposure to the optical rearrangement. Thus, a procedure similar to the no prism awareness condition was used to adjust the prisms prior to the exposure period, and a procedure similar to the complete prism awareness condition was used to adjust the prisms subsequent to the exposure period.

In addition to the three different levels of prism awareness, three different conditions of exposure, or training procedures, were employed. For *S*s participating in the concurrent display condition *E* removed a panel in the horizontal surface of the apparatus that allowed *S* to view his hand and arm during exposure. The *S*'s limb was strapped to a pivot board mounted on the table. This pivot board rotated in a 25° arc in the horizontal plane. The *S* was instructed to watch his right index finger as he moved his arm back and forth in rhythm to an audible click occurring every 3 sec. The *S*'s index finger was in the field of view during the entire exposure period. There were no explicit targets to localize in the concurrent display condition; *S* merely viewed his finger while he moved it back and forth 30 times. All *S*s were allowed to rest for 30 sec. halfway through the 30 trials of all three exposure conditions.

In the terminal display condition *S* localized the five stimulus lines labeled A, B, C, D, and E. The *S* was to reach out and point to the line corresponding to the letter designated on a prerecorded tape and then return his hand to the starting position on the table. The horizontal distance from the starting pad to the target was 38 cm. A letter was called out every 3 sec., and each of the five targets was designated an equal number of times in a randomly determined order. For the terminal display condition a special panel was installed in the opening in the top surface of the apparatus. This panel permitted *S* to view only the stylus only when his arm was extended and touching the vertical panel in the area of the exposure targets on the stimulus display panel. The *S* was instructed not to make corrections should the target be missed, but rather to try to be more accurate on subsequent trials. Thus, *S*s in this condition received visual feedback regarding localizing movements only after the movement had been initiated and was near the terminus of excursion toward the target.

In the cognitive feedback condition *S* localized the same set of stimulus targets employed in the terminal display. However, in this condition the entire top surface of the apparatus remained covered as it was during the pre- and postexposure periods. Thus, *S* received no visual feedback regarding target localizations made during the exposure period. Rather, *E* informed *S* verbally how close the stylus was to the designated target line. The verbal feedback was given at the terminus of a localization to the nearest ¼ in. left or right.

Following the exposure period *E* adjusted the prism back to 0 diopters according to the appropriate prism awareness condition. The apparatus lights were turned off and the full horizontal panel was in place for all groups during the subsequent postexposure measures. The order and presentation conditions were the same as those employed during the preexposure measures.

In addition to the nine different experimental conditions there were two control conditions. The pre- and postexposure measures were obtained under identical conditions for the experimental and control groups. The *S*s in one control condition (I) made pointing localizations of targets with no optical displacement during the exposure period. The apparatus was in the same configuration as was used for the cognitive feedback condition during exposure. Half of the *S*s received verbal error feedback in this control condition and the other half received no feedback regarding the accuracy of their pointing localizations. Thus, any shifts in localizations produced solely as a result of making pointing localizations, with or without feedback, would be manifested by these subgroups in this condition. In the other control condition (II) *S*s made only visual localizations of targets with optical displacement; no pointing localizations were made during the exposure period. Again the apparatus was in the configuration used for cognitive feedback, and *S*s merely looked at the appropriate target as its letter was designated on the prerecorded tape. For half of the *S*s in this condition displacement was to the left and for the other half it was to the right. Thus, any shifts in localizations produced by viewing targets through a displacing prism would be manifested in Control Condition II.

RESULTS

The data of primary interest in this study came from preexposure to postexposure

TABLE 1

Mean Adaptation Scores (in degrees)

Exposure condition	Prism awareness[a]	Measure[b]		
		VM	PM	VPM
Terminal display with target	NPA	4.1*** (±2.9)	1.4 (±4.9)	6.1**** (±1.7)
	PPA	1.9** (±1.7)	2.1*** (±1.6)	4.4**** (±1.0)
	CPA	2.0**** (±1.2)	2.2 (±4.9)	3.3**** (±1.3)
	Overall	2.7**** (±1.1)	1.9 (±2.1)	4.6**** (± .8)
Cognitive feedback with target	NPA	1.8* (±2.3)	1.7 (±2.7)	4.5**** (±2.7)
	PPA	1.4 (±2.2)	− .6 (±2.6)	1.4** (±1.3)
	CPA	2.8**** (±1.4)	1.8 (±4.9)	1.1 (±2.0)
	Overall	2.0**** (±1.0)	1.0 (±1.9)	2.3**** (±1.2)
Concurrent display without target	NPA	− .4 (±2.1)	2.7** (±2.3)	1.8* (±1.5)
	PPA	.3 (±2.3)	3.0** (±2.9)	.3 (±1.7)
	CPA	.8 (±1.5)	1.2 (±2.6)	.5 (±1.7)
	Overall	.2 (±1.0)	2.3**** (±1.4)	.9* (± .9)
Control I. Pointing–no displacement		− .4 (±1.0)	1.0 (±1.9)	.3 (±1.1)
Control II. No pointing–displacement		.6 (±1.8)	.2 (±2.0)	.1 (±1.0)

Note. Numbers in parentheses are 95% confidence intervals. The data presented in this table were pooled over feedback conditions for Control I and over prism orientations for all other conditions.

[a] Levels of prism awareness: NPA = no prism awareness; PPA = partial prism awareness; and CPA = complete prism awareness.

[b] Measures: VM = visuomotor; PM = proprioceptive-motor; and VPM = visuo-proprioceptive-motor.

*$p < .05$.

**$p < .025$.

***$p < .01$.

****$p < .005$.

difference scores on the VM, PM, and VPM measures. For conditions involving prismatic displacement these difference scores were signed such that positive entries indicated that the shifts in mean localization were in the direction compensating for prism-induced error. For the control condition in which there was no prismatic displacement (I) the differences were arbitrarily signed such that positive entires represented shifts to the right and negative entries shifts to the left.

Table 1 represents these mean difference scores obtained for the various exposure conditions and levels of prism awareness for the experimental groups and similar scores for the control groups. As indicated in this table, it was found that in all but one case (concurrent display–complete prism awareness) the experimental treatment conditions produced significant adaptive shifts in localizations for at least one of the measures. In contrast, the control conditions produced no shifts in localizations that were significantly different from zero. Separate repeated-measures, randomized-block-design analyses of variance were used to test for possible differences within each control condition. In Control Condition I neither the main effects nor the interaction for feedback and measures was statistically significant. Similarly,

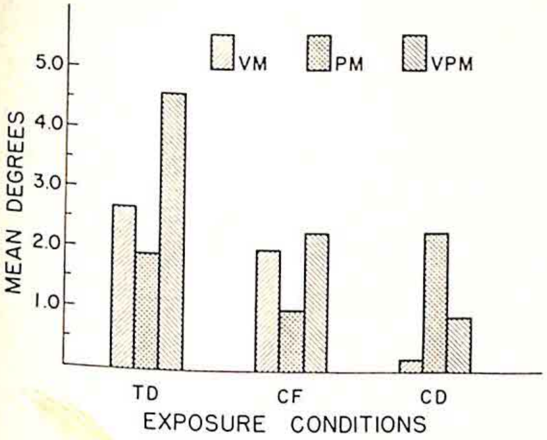


FIG. 1. Mean visuomotor (VM), proprioceptive-motor (PM), and visuo-proprioceptive-motor (VPM) shifts for terminal display (TD), cognitive feedback (CF), and concurrent display (CD) exposure conditions.

there were no significant main effects or interaction for prism orientation and measures in Control II. Prism orientation was also nonsignificant for the analysis of variance of the experimental conditions discussed below. Thus, in Table 1 the data were pooled over the two feedback conditions for Control I and prism orientations for all other conditions.

Table 2 presents the results of a repeated-measures, randomized-block-design analysis of variance of the adaptation scores for the experimental groups presented in Table 1. As indicated in Table 2 the main effect of exposure condition (A) was statistically significant, as was the interaction between exposure condition and measure (D). The significant A main effect reflects the relatively large overall shifts for the terminal display ($\bar{X} = 3.0^\circ$) compared to the relatively smaller shifts for the cognitive feedback ($\bar{X} = 1.8^\circ$) and the concurrent display ($\bar{X} = 1.1^\circ$). A breakdown that compares the combined shifts for cognitive feedback and concurrent display with the shifts for the terminal display accounted for 90% of the variation in the A main effect.

The means of the conditions represented in the A × D interaction are contained in the entries labeled "Overall" in Table 1. Figure 1 presents this interaction graphically. In Table 1 it can be seen that the VM and VPM shifts were significantly greater than zero for the terminal display and cognitive feedback conditions, while the PM shifts were nonsignificant. For the concurrent display the opposite relationship for the VM and PM measures obtained. There was also a smaller but statistically significant VPM shift for the concurrent display. A breakdown that compares the average of the terminal display and cognitive feedback conditions with the concurrent display condition for the three dependent measures accounts for 88% of the variation in the A × D interaction. This indicates that the significant variation represented in the A × D interaction was due primarily to variation in the relative magnitudes of the VM, PM, and VPM measures for the concurrent display compared with the magnitudes of these measures averaged across the terminal display and cognitive feedback conditions.

The variation in the interaction between the variables of prism awareness (B) and measure (D) was not statistically significant. However, it can be seen in Figure 2 that most of the variation in this interaction was in the values of the VPM measures for the different levels of prism awareness. The A × B × D interaction in Table 2 was also not statistically significant.

TABLE 2

ANALYSIS OF VARIANCE OF MEAN ADAPTATION SCORES FOR EXPERIMENTAL CONDITIONS

Source of variation	*df*	*MS*	*F*
Exposure condition (A)	2	67.49	7.17**
Prism awareness (B)	2	23.42	2.49
Prism orientation (C)	1	9.80	1.04
A × B	4	7.24	<1
A × C	2	21.22	2.26
B × C	2	3.33	<1
A × B × C	4	26.71	2.84
Error (a), *S*s (ABC)	54	9.41	
Measure (D)	2	19.70	1.73
A × D	4	32.78	2.88*
B × D	4	12.52	1.10
C × D	2	0.48	<1
A × B × D	8	7.55	<1
A × C × D	4	2.29	<1
B × C × D	4	3.39	<1
A × B × C × D	8	7.35	<1
Error (b), *S*s (ABC)D	108	11.37	

* $p < .05$.
** $p < .005$.

The means of the individual cells contributing to this interaction are presented in Table 1. Examination of these data shows systematic variation in the VPM measure both as a function of prism awareness and exposure condition. Figure 3 presents the VPM data graphically. A separate analysis of variance was used to analyze this VPM variation. The main effect for exposure condition was statistically significant, F (2, 63) = 17.23, $p < .005$, indicating significant variation between the terminal display ($\bar{X}$ = 4.6°), cognitive feedback ($\bar{X}$ = 2.3°), and concurrent display ($\bar{X}$ = .9°). The main effect of prism awareness was also statistically significant, F (2, 63) = 9.11, $p < .005$, indicating significant variation between no prism awareness ($\bar{X}$ = 4.1°), partial prism awareness ($\bar{X}$ = 2.0°), and complete prism awareness ($\bar{X}$ = 1.6°). Multiple comparisons of the treatment means for both main effects were made using Duncan's range test, R_3 (63) = 1.3°, R_2 (63) = 1.3°, $p < .05$ (cf. Edwards, 1972). The interaction between exposure condition and prism awareness was nonsignificant in this VPM analysis.

Discussion

In general the data provide support for the hypotheses derived from a general information-processing approach to prism adaptation. The finding of significant adaptation for all three types of training procedures indicates that qualitatively different kinds of information serve an equivalent function, viz., reduction of the discordance and hence a more accurate ability for spatial orientation. According to an information-processing formulation the condition necessary to produce adaptation is some sort of information regarding the altered state of the system. Such information was the factor in common for all of the conditions that evidenced significant adaptive shifts in localizations. However, it appears that the qualitatively different kinds of information are only functionally equivalent in the sense that they were sufficient to produce adaptive shifts. The actual locus and nature of these shifts varied as a function of the source of relevant information.

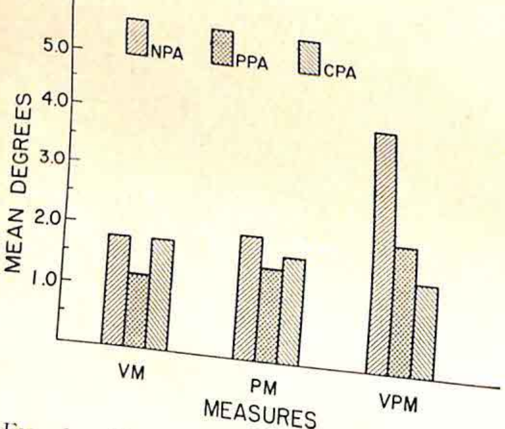


FIG. 2. Mean visuomotor (VM), proprioceptive-motor (PM), and visuo-proprioceptive-motor (VPM) shifts for no prism awareness (NPA), partial prism awareness (PPA), and complete prism awareness (CPA) conditions.

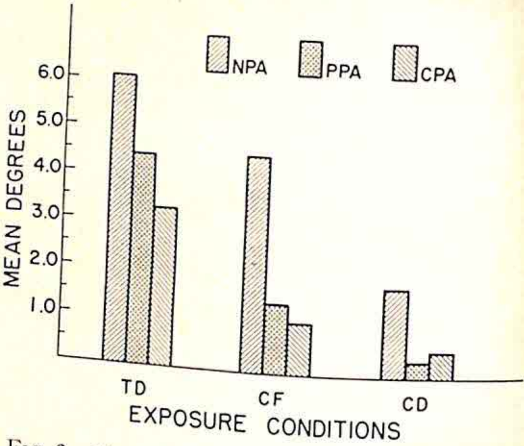


FIG. 3. Mean visuo-proprioceptive-motor (VPM) shifts for the three exposure conditions (TD = terminal display; CF = cognitive feedback; CD = concurrent display) and three levels of prism awareness (NPA = no prism awareness; PPA = partial prism awareness; CPA = complete prism awareness).

The theory proposed by Canon and Uhlarik (Canon, 1970, 1971; Uhlarik & Canon, 1971) makes specific predictions regarding the relationship between the source of relevant information and the loci of the adaptive shifts in localizations. The model predicts that the recalibration of discordant sensorimotor inputs will manifest itself in the modality *not* utilized as a source of information for localizing stimulus objects. In accordance with this model it was demonstrated (Uhlarik & Canon, 1971) that the physical conditions of exposure to rearrangement produce differences in the modality in which adaptation will be manifested. The results of the present study were consistent with the model. As hypothesized, there was a significant interaction between the condition of exposure and the nature of the adaptive shifts as indexed by the VM and PM measures. In both the terminal display and the cognitive feedback conditions, visual feedback from Ss' pointing limbs was precluded as a basis for localizing the targets. Hence, Ss in these conditions were forced to attend to the relationship between the PM system and the locus of the target. In both of these conditions there were relatively large and statistically significant overall VM shifts and smaller PM shifts that were generally not statistically greater than zero. On the other hand, even though there were no explicit visual targets to localize in the concurrent display condition, the model predicts the opposite relationship for the VM and PM measures. In view of the evidence on visual dominance when looking through a wedge prism (Gibson, 1933; Hay, Pick, & Ikeda, 1965; Rock, 1966), it can be assumed that Ss in this condition would be attending primarily to the visual impression of their limbs. Thus, the PM system would be the nonattended system that should evidence the adaptive shifts. The data from the concurrent display support this prediction. There were generally relatively large and statistically significant PM shifts and smaller VM shifts that were not statistically greater than zero.

The finding of large VM shifts and small PM shifts with the terminal display confirms previous findings (Uhlarik & Canon, 1971). The present results follow a pattern of performance on the VM and PM measures for the concurrent display with no targets which was essentially the same as obtained in previous research using a concurrent display with targets (Uhlarik & Canon, 1971). Thus, the generality of the formulation is extended, at least for conditions of the concurrent display, to situations in which there are no targets to localize and no other source of error information. Of greater interest, however, is the finding that cognitive feedback condition was sufficient to produce adaptation in the absence of reafferent feedback that Held (1968) has hypothesized to be essential for genuine adaptation. The presence of both significant overall VM and VPM shifts in the cognitive feedback condition imply that there was in fact "genuine" sensorimotor adaptation and that Ss were not merely applying some sort of conscious correction to their localizations. A decision rule to point consistently to one side of the targets would be an example of such conscious correction that would manifest itself primarily on measures that involve pointing at visual targets (e.g., VPM measure). Presumably the VM and PM measures would be less prone to such conscious manipulation. In the case of VM adaptation, for example, an S would have to know that adaptive VM shifts would be in the direction opposite that of adaptive VPM shifts. Thus, the findings that, for the cognitive feedback condition, adaptation was primarily VM in nature militates against an interpretation that adaptation was due to conscious correction during the periods in which the measures were obtained. Because Ss have to attend to the PM system in the cognitive feedback condition the information-processing formulation presented above predicts that the adaptation would be primarily VM rather than PM in nature.

The present data also support the prediction that the overall magnitude of the adaptive shifts would vary as a function of the amount of available information. The terminal display, containing both cognitive information in the form of target-error feedback and information regarding intermodality discordance, produced overall adaptive shifts of greater magnitude than did either the concurrent display or cognitive feedback conditions. The concurrent display, which primarily contained information regarding intermodality discordance, produced overall adaptive shifts that were slightly smaller in magnitude than those produced by the cognitive feedback condition, which contained primarily target-error information. However, it should be noted that there was confounding between the number of movements made during exposure and duration of the movements. The experiment was designed to control exactly the number of pointing localizations made during exposure to the terminal display and cognitive feedback with the back-and-forth transverse movements made during exposure to the concurrent display. Movement with the terminal and concurrent displays was paced to occur every 3 sec. For cognitive feedback the movement was not rhythmically paced, due to the fact that E had to measure localization errors and provide this feedback to S verbally. Each localization movement required an average duration of approximately 5 sec. for the cognitive feedback condition. Thus, overall differences among exposure conditions could not be unambiguously interpreted, because variation of kind and amount of available information was confounded with duration of and type of movement made during the exposure periods.

The variable of prism awareness produced systematic variation in the magnitude of the VPM shifts. It is important to note that this variable did not significantly alter the magnitudes of the VM and PM shifts. This implies that the effects of explicit information regarding the optical effects of the prism alter only that measure that is assumed to be most prone to systematic conscious correction. Generally for all exposure conditions, no prism awareness produced the largest VPM shifts and complete prism awareness produced the smallest VPM shifts. Similarly the terminal display generally produced the largest VPM shifts and the concurrent display generally produced the smallest VPM shifts. Aftereffects such as the VPM measure have been typically viewed as the manifestation of the persistence of the adaptive sensorimotor adjustment that develops during exposure to optical rearrangement. The assumption is that such dependent measures are the complement of adaptation and accurately index the process underlying adaptation. Previous research has demonstrated that the complementary relationship between measures of adaptation and aftereffect does not always obtain (e.g., Uhlarik & Canon, 1970; Wallach, Kravitz, & Lindauer, 1963). It has never been demonstrated, however, whether such variations are due to alterations in the actual process underlying adaptation or in the indirect measures of that process. The results of the present study suggest that the effect of explicit prism awareness was to produce variation in the most commonly used measure of the adaptive process. That is to say, the VPM measure appears to be the one affected while the PM and VM measures remain unchanged by this manipulation. These data indicate reason for methodological caution because it appears possible to manipulate the magnitude of VPM measures independently of alteration of the amount of underlying adaptation as indexed by other measures.

REFERENCES

CANON, L. K. Intermodality inconsistency of input and directed attention as determinants of the nature of adaptation. *Journal of Experimental Psychology*, 1970, **84**, 141–147.

CANON, L. K. Directed attention and maladaptive "adaptation" to displacement of the visual field. *Journal of Experimental Psychology*, 1971, **88**, 403–408.

EDWARDS, A. L. *Experimental design in psychological research.* (4th ed.) New York: Holt, Rinehart & Winston, 1972.

FREEDMAN, S. J. On the mechanisms of perceptual compensation. In S. J. Freedman (Ed.), *The neuropsychology of spatially oriented behavior.* Homewood, Ill.: Dorsey Press, 1968.

GIBSON, J. G. Adaptation, aftereffects, and contrast in the perception of curved lines. *Journal of Experimental Psychology*, 1933, **16**, 1–31.

HARDT, M. E., HELD, R., & STEINBACH, M. J. Adaptation to displaced vision: A change in the central control of sensorimotor coordination. *Journal of Experimental Psychology*, 1971, **89**, 229–239.

HAY, J. C., PICK, H. L., & IKEDA, K. Visual capture produced by prism spectacles. *Psychonomic Science*, 1965, **2**, 215–216.

HELD, R. Plasticity in sensorimotor coordination. In S. J. Freedman (Ed.), *The neuropsychology of spatially oriented behavior.* Homewood, Ill.: Dorsey Press, 1968.

HELD, R., & HEIN, A. V. Adaptation of disarranged hand-eye coordination contingent upon re-afferent stimulation. *Perceptual and Motor Skills*, 1958, **8**, 87–90.

HOWARD, I. P. Displacing the optical array. In S. J. Freedman (Ed.), *The neuropsychology of spatially oriented behavior.* Homewood, Ill.: Dorsey Press, 1968.

HOWARD, I. P. The adaptability of the visual-motor system. In K. J. Connoly (Ed.), *Mechanisms of motor skill development.* London: Academic Press, 1970.

ROCK, I. *The nature of perceptual adaptation.* New York: Basic Books Inc., 1966.

UHLARIK, J. J. A device for presenting targets and recording positioning responses in one dimension. *Behavior Research Methods and Instrumentation*, 1972, **1**, 15–16.

UHLARIK, J. J., & CANON, L. K. Effects of situational cues on prism-induced aftereffects. *Perception & Psychophysics*, 1970, **7**, 348–350.

UHLARIK, J. J., & CANON, L. K. Influence of concurrent and terminal exposure conditions on the nature of perceptual adaptation. *Journal of Experimental Psychology*, 1971, **91**, 233–239.

WALLACH, H., KRAVITZ, J. H., & LINDAUER, J. A passive condition for rapid adaptation to displaced visual direction. *American Journal of Psychology*, 1963, **76**, 568–578.

WOHLWILL, J. F. Perceptual learning. In P. R. Farnsworth (Ed.), *Annual review of psychology.* Palo Alto, Calif.: Annual Reviews Inc., 1966.

(Received August 7, 1972)

Journal of Experimental Psychology
1973, Vol. 98, No. 2, 233–238

PROACTIVE EFFECTS OF INTERPOLATED ANCHORS[1]

HADYN D. ELLIS[2]

University of Aberdeen, Aberdeen, Scotland

An experiment was performed in which eight *S*s made comparative judgments of two objectively equally loud stimuli. Louder and softer anchor stimuli occurred before, between, or after the comparison pair in order to decide whether anchors operate proactively (i.e., producing an effect on the reception of the *subsequent* stimulus) or retroactively (i.e., affecting the memory trace of the *preceding* stimulus). The results are interpreted to mean that anchors act proactively. Two subsidiary investigations were also carried out. In the first, order effects were examined. These sometimes arise as the result of making comparisons of the first vs. the second stimulus as opposed to the second vs. the first. No such order effect was found. The second investigation concerned the effect of the duration of time elapsing between the comparison stimuli with no anchor stimulus present. With an interstimulus interval of $3\frac{1}{2}$ sec., a pronounced negative time error was found (i.e., the tendency to judge the second of two equal stimuli to be louder than the first). With an interval of $1\frac{1}{2}$ sec., no such error was observed.

When applying the principles of adaptation-level theory (ALT) to comparative judgments involving interpolated anchors, Helson (1964) makes the implicit assumption that such anchors pool with the standard stimulus (St) as well as previous stimuli to form an adaptation level (AL) against which the variable stimulus (Va) is judged. That is, he envisages a process of *retroactive assimilation*.

This model certainly fits the published data (e.g., Guilford & Park, 1931; Needham, 1935; Philip, 1947; Pratt, 1933), wherein anchors greater than the St and the Va produce a tendency to underestimate the latter, while lesser anchors generally lead to its overestimation. In Helson's (1964) view the AL would be, respectively, high and low.

These results, though, could just as well be accounted for by a model involving the process of *proactive contrast* between the anchor and the Va. This model ignores any changes in the memory of the St, and instead proposes that greater anchors act to depress the sensation of the subsequent Va. Lesser anchors, on the other hand, lead to an enhancement of the apparent intensity of the Va.

In order to test this alternative model, a comparative judgment technique was used that has been described in detail elsewhere (Ellis, 1972). Basically, it involves the comparison of a "standard" stimulus (K_1) with itself (K_2) while an anchor stimulus (A) is interpolated between them. This design, known as the KAK paradigm, is chosen on the grounds of expediency. The *S*s typically fail to realize that K_1 and K_2 are equal since the presence of the interpolated anchor usually produces the illusion that $K_2 > K_1$, or $K_2 < K_1$, depending on whether the anchor is lesser or greater than the stimulus. Thus far, the two models under consideration are equally plausible, since both predict the same results when anchors are interpolated between K_1 and K_2.

When, however, anchors are extrapolated instead of being interpolated between K_1 and K_2, the two predict quite different outcomes. According to Helson's (1964) model, placing the anchor before K_1 should not markedly alter the effect produced when it follows K_1, since in both cases they will pool to form the AL against which K_2 is judged. On the basis of proactive contrast, however, the two designs should result in opposite outcomes: In the

---

[1] This experiment was carried out while the author held a Science Research Council Scholarship at Reading University, Reading, England, under the aegis of M. Treisman.

[2] Requests for reprints should be sent to H. D. Ellis, Department of Psychology, University of Aberdeen, King's College, Old Aberdeen, AB9 2UB, Scotland, United Kingdom.

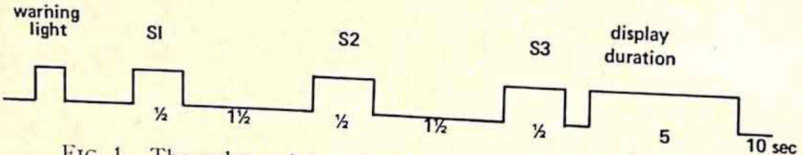


FIG. 1. The order and duration (in sec.) of events in each trial.

one instance the anchor will operate proactively on K_1, and in the other on K_2. When the anchor follows K_2, though, there should be no discernible effect according to the proactive contrast argument, but, on the basis of retroactive assimilation (ALT, to be sure, is not clear here), the anchor, in pooling with K_2, should produce the opposite effect to when it is interpolated.

In addition to testing these models, the following experiment aims to discover whether response order effects occur in the KAK situation, i.e., whether making judgments of K_2 relative to K_1 produces different results from judgments of K_1 relative to K_2. Such a difference might be expected from the comparative judgment literature wherein standard vs. variable judgments have sometimes been reported to differ from variable vs. standard judgments (e.g., Michels & Helson, 1954).

METHOD

Subjects. Four male and four female *S*s (aged 18–24 yr.) were paid to take part in two sessions.

Apparatus. The stimuli were 900-Hz. tones generated by an Advance Components, Ltd. (Jz) oscillator fed through a bank of Advance Components, Ltd. attenuators and delivered binarually via Dynamic Associates earphones to *S*, who was seated in an Industrial Acoustics Corp. double soundproof chamber. Stimulus durations and interstimulus intervals were controlled by an AIM bio Sciences, Ltd. system of timers, switches, and relays. Their onset and offset were gradual (20 msec. and 80 msec., respectively), in order to avoid clicks. The duration of events within each trial is illustrated in Figure 1.

In front of *S* was a panel containing a warning light which heralded each new trial and two digital display tubes that served to cue him as to which two of the three tones presented on each trial to compare in terms of loudness.

Procedure. When *S* was comfortably seated in the chamber he was given the following instructions:

> The following experiment is an investigation of how people make relative judgments of loudness when the differences between stimuli are small. Each trial is preceded by a flash from the red warning light in front of you and usually comprises three tones from which you must compare two. You will, however, be ignorant of which two to compare until you have heard all three (or sometimes only two will be presented). Whether you receive two or three tones, the last will be followed by ignition of the two digital display tubes [show]. This will read, for example, "2 vs. 1," which means that you must recall the second tone and compare it with the first tone. You must respond either "louder" or "softer" as promptly as possible. No "equal" judgments are allowed so if you really can't decide, then guess.

Following this, *S*'s threshold for a continuous tone was assessed by the method of limits and K_1 and K_2 set at 60 db. above this value.

Fifty practice trials were then given in order to familiarize *S* with the rather complex procedure. Ten of these trials involved real differences so that *E* could note whether *S* had failed to grasp the procedure. In fact, no *S* required remedial training. After the practice trials *S*s were allowed a 5-min. break before proceeding with the first session. The second session (held on a subsequent day) was preceded by threshold establishment and a few practice trials. The first session took 80 min. and the second took about 60 min. for each *S*. In both sessions a 5-min. break was given at the halfway stage. Table 1 summarizes the distribution of the 114 trials for the five conditions experienced by each *S*, i.e., anchors preceding (P), interpolated (I), and following (F) the comparison stimuli, as well as two no-anchor conditions.

The three anchors used were 40, 60, and 80 db. (SL) and each of these occurred a total of 60 times (30 each session). The three anchor positions were each used a total of 20 times, and the six possible number combinations (i.e., 3 vs. 1, 3 vs. 2, 2 vs. 1, 2 vs. 3, 1 vs. 3, and 1 vs. 2) each occurred a total of 30 times. These were all separately randomized for individual *S*s.

TABLE 1

DISTRIBUTION OF TRIALS AMONG THE FIVE CONDITIONS

No. of trials	Condition	Comparisons
30	Preceding (P) = AKK	3 vs. 2; 2 vs. 3
30	Interpolated (I) = KAK	3 vs. 1; 1 vs. 3
30	Following (F) = KKA	2 vs. 1; 1 vs. 2
12	KK($3\frac{1}{2}$-sec. interval)	2 vs. 1; 1 vs. 2
12	KK($1\frac{1}{2}$-sec. interval)	2 vs. 1; 1 vs. 2

Note. K = standard stimulus; A = anchor.

There were 24 further trials (12 per session) in which only K_1 and K_2 were presented; half of these involved an ISI (interstimulus interval) of $3\frac{1}{2}$ sec., and the remainder, an ISI of $1\frac{1}{2}$ sec.

RESULTS

The proportion of judgments $K_2 < K_1$ for all conditions, each anchor, and both orders of judgment are given in Table 2 and illustrated in Figure 2.

TABLE 2

PROPORTION OF JUDGMENTS $K_2 < K_1$ FOR EACH *S* UNDER EACH CONDITION

S	Order of judgment					
	K_2 vs. K_1			K_1 vs. K_2		
	Anchor (db. SL)			Anchor (db. SL)		
	40	60	80	40	60	80
	Condition P (preceding)					
1	.70	.50	.50	.90	.50	.40
2	.40	.60	.70	.70	.80	.50
3	.60	0.00	.10	.60	.90	.20
4	0.70	.60	.80	.70	.50	.50
5	.90	.90	.20	1.00	.70	.20
6	.20	.60	.60	.50	.20	.20
7	.70	.80	.60	.80	.70	.60
8	.70	.50	.50	.50	.30	.40
$\bar{X}$	.61	.56	.50	.71	.58	.38
	Condition I (interpolated)					
1	.40	.80	1.00	.90	.80	1.00
2	.30	.60	.70	.30	.40	.60
3	.10	0.00	.70	.30	.20	.50
4	0.00	.60	.80	.10	.70	.90
5	.70	1.00	.70	.60	.70	.50
6	.10	.40	.40	.60	.60	.80
7	.10	.80	1.00	.30	.80	.90
8	.20	.50	.30	.40	.30	.80
$\bar{X}$	.24	.59	.70	.44	.56	.75
	Condition F (following)					
1	.90	.50	.70	.70	.90	.50
2	.30	.40	.10	.20	.40	.20
3	.40	.20	.20	.50	.40	.20
4	.30	.20	.70	.40	.50	.40
5	.70	.40	.60	.60	.60	.30
6	.50	.50	.70	.70	1.00	.50
7	.70	1.00	.60	.90	.80	.80
8	.60	.60	.50	.20	.40	.30
$\bar{X}$	.55	.48	.51	.53	.63	.40

Note. In order to make the K_1 vs. K_2 judgments compatible with the K_2 vs. K_1 judgments they were converted to proportions and this value subtracted from unity. (Thus .1 judgments $K_1 < K_2$ are equivalent to a proportion of .9 $K_2 < K_1$ judgments.)

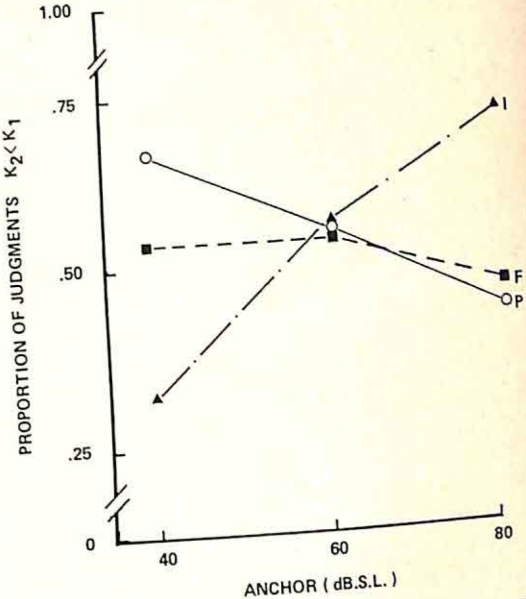


FIG. 2. The proportion of judgments $K_2 < K_1$ at each anchor level (collapsed across order conditions) for the three conditions: P (preceding), I (interpolated), and F (following).

The data were submitted to an analysis of variance, which indicated a significant Anchors × Conditions interaction, F (4, 28) = 15.28, $p < .001$. None of the interactions involving the order-of-judgment factor were significant, so the data were collapsed and each condition analyzed for anchor effects by an *S*s × Treatments analysis. Both Condition P and Condition I yielded significant anchor effects, F (2, 14) = 4.1 and 10.4, respectively; whereas Condition F produced an $F = 1.5$, which is not significant.

Condition P differed from Condition I both at the 40-db. anchor, t (7) = 4.6, $p < .01$, and the 80-db. anchor, t (7) = 4.4., $p < .01$. The P and I curves were further analyzed by the Newman-Keuls method in order to determine differences between anchor treatments. This indicated that, for Condition P, the 40-db. anchor scores differed from those of the 80-db. anchor ($p < .05$), but the 40–60 db. and 80–60 db. comparisons yielded no significant differences. For Condition I, it was found (again with a criterion of $p = .05$) that the

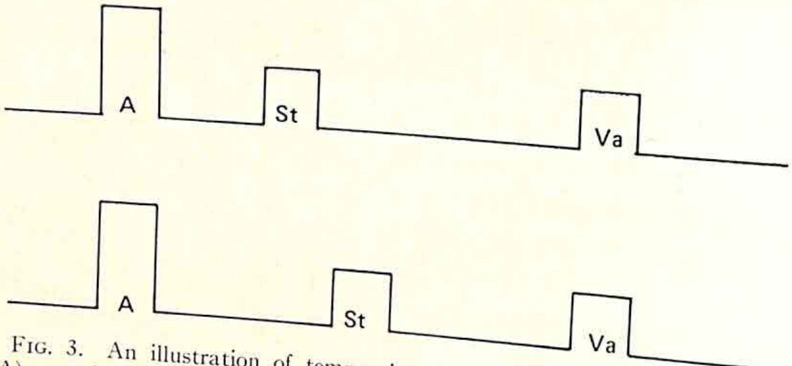


FIG. 3. An illustration of temporal asymmetry (top) when the anchor (A) precedes the standard (St) and variable (Va) stimuli, compared with the symmetry (bottom) employed in the present study.

40-db. anchor scores differed significantly from both the 60-db. and 80-db. anchor scores, but that the 80-db. treatment did not differ from that of the 60-db. anchor. Thus anchor effects are more reliable in Condition I than Condition P, and absent in Condition F.

Results of trials without an anchor stimulus. There was no appreciable order effect (i.e., K_1 vs. K_2 or K_2 vs. K_1) and so the results were pooled.

The longer ISI showed a marked negative time error in that the second of the two stimuli was judged louder than the first an average of 83% of the time. With an ISI of $1\frac{1}{2}$ sec., this trend was hardly visible, for .53 of the judgments were $K_2 > K_1$, which is quite close to a chance proportion.

The two conditions were compared by a nonparametric test (Wilcoxon matched pairs) because the distribution of scores in the ISI $3\frac{1}{2}$-sec. condition was negatively skewed. This indicated a significant difference, $T\ (8) = 1.5$, $p < .02$. The mean for an ISI of $1\frac{1}{2}$ sec., in fact, did not significantly differ from chance ($t < 1$), and so it can be concluded that no time error was found for that interval.

DISCUSSION

Direction of anchor effect. The data suggest that the effects of interpolated anchors are the result of a process of proactive contrast rather than one involving retroactive assimilation. This must be qualified by the following considerations.

First, the experimental procedure appears to have slightly weakened the KAK effect (Condition I): There is not such a distinct difference among anchors as has been found in other studies (Ellis, 1971). Normally the 60- and 80-db. anchors produce significantly different proportions of judgments $K_2 < K_1$. Such a diminution of effect could conceivably indicate that the present procedure is not typical of the usual interpolated design, and so any conclusions drawn from those instances where anchors were extrapolated might be spurious. The evidence, though, points to a quantitative rather than a qualitative change, which might easily have resulted from using fewer trials per condition than usual with a relatively small sample of *S*s.

Second, this conclusion is at variance with at least four experimental studies as well as a great deal of theory concerning the effects of interpolated stimuli upon time errors (e.g., Pratt, 1936).

The studies alluded to were also aimed at elucidating time error (Gleitman, 1957; Gleitman & Hay, 1964; Kind & Brown, 1966; Philip, 1947). In each case, preceding and interpolated anchors were used. (Only Philip's study employed an anchor following the comparison stimuli.) All of the studies found evidence suggesting that assimilation between the St and anchor mediates judgmental distortions. Apart from Philip's study, however, these findings could be interpreted in terms of time error, since only anchors greater than the comparison stimuli were used. That is, it is conceivable that when preceding, anchors actually depressed the apparent value of the St, but this process may have been masked by strong negative time errors. The use of equal and lesser anchors permits some evaluation of this hypothesis. In the present study the

60-db. anchor produced similar judgments in both P and I conditions. The relationship between time error and interpolated anchor effects is discussed in detail elsewhere (Ellis, 1971). It should be noted, of course, that Philip did employ an equal and a lesser anchor. His results might be explained by the fact that he used temporal stimuli, which need not be subject to the same laws as those governing sensory stimuli. Indeed, there is evidence to suggest that time estimations are subject to assimilation rather than contrast (Treisman, 1963).

The final point to be made concerning the discrepancy between the findings of the above four investigations and the present one is that each involved the temporal asymmetry illustrated in Figure 3. In order to maintain constant St − Va and A − St (or St − A) intervals, there resulted a sort of grouping between the anchor and St that might itself effect assimilation. Such grouping is absent in the usual interpolated design and avoided in the present study at the expense of a varying $K_1 - K_2$ interval, but this should only act to increase negative time error (Ellis, 1971), which will facilitate some anchors and diminish the effects of opposite anchors. Whichever are enhanced or diminished will depend upon whether Condition P or Condition I is used. With the above considerations in mind, it would seem that there is a possibility that interpolated anchors act proactively to cause a contrast effect upon the perception of a subsequent stimulus. This undoubtedly corresponds with the more usual manner of studying context effects. This typically involves category scaling with an anchor being presented *before* each series stimulus.

Additionally, recent electrophysiological evidence indicates that averaged auditory evoked responses to a stimulus recorded from the scalp are larger or smaller depending on whether the preceding anchor stimulus is smaller or larger (Buchsbaum, Silverman, Henkin, & Pfefferbaum, 1971). Given that such evoked responses are said to positively correlate with the dimension of loudness (Davis & Zerlin, 1965), this seems to be reasonable, independent evidence for the operation of proactive contrast.

Order of responses. The finding that it makes no difference whether judgments are made K_1 relative to K_2 or K_2 to K_1 is interesting in that it suggests that postcuing as to which stimulus to consider as the St eradicates the so-called order error which has frequently been reported. Michels and Helson (1954), in fact, offer different ALT formulations depending upon the order of judgment. The present data suggest that the error is a methodological artifact that is easily circumvented.

No-anchor trials. The trials in which only K_1 and K_2 appeared indicated that a strong negative time error was operating. With an interval of $3\frac{1}{2}$ sec. there was a preponderance of $K_2 > K_1$ judgments that was absent when the ISI was only $1\frac{1}{2}$ sec. The relationship between time error and the KAK contrast effect has already been considered (Ellis, 1971). The two appear to be unrelated in that increasing the ISIs leads to a decrease in the effects of greater anchors, while it causes an increase in the effects of lesser anchors. This would be expected on the basis of negative time error, which would tend to produce more $K_2 > K_1$ judgments.

The data from those trials with no anchor serve to demonstrate that when $A = K_1 = K_2$ the time error is not equivalent to that when K_1 and K_2 are separated by the same time interval without an interpolated anchor. In the former case the probability of a judgment $K_2 > K_1$ is .43; in the latter it is .83. It is possible that *S*s use the anchor equal to K_1 (i.e., the 50-db. anchor) as a reference for K_1. If this is the case, the effective ISI will be $1\frac{1}{2}$ sec. rather than $3\frac{1}{2}$ sec., which would eliminate the negative time error bias.

REFERENCES

Buchsbaum, M., Silverman, J., Henkin, R. I., & Pfefferbaum, A. Contrast effects on the auditory evoked response and its relation to psychophysical judgments. *Perception & Psychophysics*, 1971, **9**, 379–384.

Davis, H., & Zerlin, S. Acoustic relations of the human vertex potential. *Journal of the Acoustical Society of America*, 1965, **39**, 109–116.

Ellis, H. D. Context effects in sensory judgment. Unpublished doctoral dissertation, University of Reading, England, 1971.

Ellis, H. D. Anchor effects in comparative judgment and an analysis of response times. *Acta Psychologica*, 1972, **36**, 85–94.

Gleitman, H. Proactive and retroactive assimilation in the successive comparison of loudness. *American Journal of Psychology*, 1957, **70**, 117–119.

Gleitman, H., & Hay, J. C. The effect of instruction on positive time errors in successive comparison. *American Journal of Psychology*, 1964, **77**, 286–290.

Guilford, J. P., & Park, D. G. The effect of interpolated weights upon comparative judgments. *American Journal of Psychology*, 1931, **43**, 589–599.

HELSON, H. *Adaptation-level theory.* New York: Harper & Row, 1964.

KIND, B., & BROWN, D. R. Time error and the separation of cognitive systems. *Journal of Personality*, 1966, **34**, 569–576.

MICHELS, W. C., & HELSON, H. A quantitative theory of time-order effects. *American Journal of Psychology*, 1954, **67**, 327–334.

NEEDHAM, J. G. Interpolation effects with different time intervals. *Journal of Experimental Psychology*, 1935, **18**, 767–773.

PHILIP, B. R. The effect of interpolated and extrapolated stimuli upon the time-order error in the comparison of temporal intervals. *Journal of General Psychology*, 1947, **36**, 173–187.

PRATT, C. C. The time error in psychophysical judgments. *American Journal of Psychology*, 1933, **35**, 292–297.

PRATT, C. C. The law of disuse. *Psychological Review*, 1936, **43**, 83–93.

TREISMAN, M. Temporal discrimination and the indifference interval: Implications for a model of the "internal clock." *Psychological Monographs*, 1963, **77** (13, Whole No. 576).

(Received June 15, 1972)

Journal of Experimental Psychology
1973, Vol. 98, No. 2, 239-245

EFFECTS OF SET-INDUCING INSTRUCTIONS ON RECALL FROM DICHOTIC INPUTS[1]

DAVID SHINAR AND MARI RIESS JONES[2]

Ohio State University

Dichotic listening strategies of *S*s were manipulated in both selective- and divided-attention situations by means of instructions. Half of the *S*s were told to expect to recall three digits from shadowed messages of variable lengths, and half were told they would be tested for digit recall primarily on the simultaneously presented nonshadowed digit string messages. The predicted Instructions × Recall Channel interaction was significant for *S*s in both selective- and divided-attention conditions. Overall, digit recall accuracy did not differ for the two instructional groups, suggesting that *S*s allot a limited amount to attentional capacity in a manner determined by instructions and task demands. Results are discussed in terms of active vs. passive theories of attention.

One feature of Broadbent's (1958) original model has remained undisputed: We are limited in our capacity to process information. Models that have attempted to account for this limit can be broadly dichotomized into passive filter vs. active nonfilter theories.

Filter theories postulate information reduction mechanisms (filters) at different points in the processing chain, and controversies tend to revolve about the locus of the filter mechanism. Treisman (e.g., 1964) posits an attenuation filter at a "perceptual" stage prior to full verbal analysis, while others (Deutsch & Deutsch, 1963; Norman, 1968) argue that selectivity precedes only response selection. Studies supporting one or the other of these models can be differentiated on the basis of experimental task and dependent measure. When *S*s are instructed to divide attention and the performance measure is one of detection of items in the nonshadowed channel, results support a perceptual filter model (Treisman & Geffen, 1967; Treisman & Riley, 1969). When *S*s must ignore one of the two channels and the performance measure is one of interference from that channel, then verbal content in that channel seems to affect the shadowing latency of the attended channel, implying that all input is fully analyzed prior to any filtration process (Greenwald, 1970; Lewis, 1970).

A possible resolution of such contradictory findings focuses upon the strategies *S*s must use in different types of dichotic listening tasks. Both instructions and task demands may determine *S*'s encoding strategy, as Harris and Haber (1963) and Bryden (1971) have shown in related tasks. Active attention models (e.g., Hochberg, 1970; Neisser, 1967) emphasize the possibility that instructions determine the allocation of attentional energy in any given task in such a manner as to optimize performance. In fact, Moray and Fitter (in press) have argued explicitly that attention should be viewed as an acquired skill whose allocation can be adjusted according "to the requirements of the task at hand."

The present study manipulated task requirements by instructions rather than the more common approach of manipulating the nature of the material (cf. Treisman, 1969). The *S*s were required to shadow one of two simultaneous dichotic messages, where both messages were random digit sequences. The *S*s were asked to recall a letter that terminated one message and then to recall the last three digits from one of the lists. The *S* always knew in advance whether the letter would be in the shadowed or the nonshadowed message, but was not informed from which list to recall the digits

---

[1] This research was supported in part by a grant to the second author from the Humanities College of Ohio State University.

[2] Requests for reprints should be sent to Mari R. Jones, Department of Psychology, Ohio State University, 404C West Seventeenth Avenue, Columbus, Ohio 43210.

before he finished shadowing the last item. Expectations for digit recall were manipulated through instructions: Half of the *S*s were told that in most trials they would be cued to recall the digits from the shadowed message and half were told they would be tested primarily on the nonshadowed message. If expectancy controls the allocation of attentional capacity, then digit recall from the shadowed message should be better for those *S*s who expect to recall from it more frequently. Similarly, recall from the nonshadowed message should be better for *S*s who expect to recall from the nonshadowed message. No difference was predicted between the two groups in terms of letter recall because all *S*s knew in advance to which ear the letter would be presented.

The first of the two predictions cannot be derived from the passive models mentioned above. Treisman's (1964) model cannot account at all for any recall (of either group of *S*s) of digits in the nonshadowed message. According to her model, information from the nonshadowed channel can be recognized only when the recognition criterion for specific items is lowered (p. 14). However the criterion cannot be lowered for an unspecified stimulus, and in the present task the to-be-remembered digits remain unspecified until the list is ended. Norman's (1968) model can account for memory of information from the nonshadowed channel but not for the difference between the two instruction groups. In his model the memory for specific items can be enhanced by giving their representation in short-term storage (STS) more "pertinence," where pertinence "must be determined before or as the input is being analyzed and selected, it cannot be determined after the selection has taken place [p. 527]." In the present study instructions were used to differentially weigh input sources, i.e., ears but not input content (i.e., specific items), and therefore the mechanism of pertinence cannot account for the superiority of one group over the other.

Finally it was hypothesized that if task requirements, as manipulated through the instructions, can alter the basic processes involved in processing and storing verbal information, then the shape of the recall functions should differ not only as a function of the message (shadowed vs. nonshadowed) but also as a function of expectancies. This prediction is contrary to a generalization of Neisser (1967, p. 213) that in dichotic presentation with shadowing the level of analysis of the two inputs is determined by their source: The nonshadowed one is stored only at an echoic level while the shadowed one is processed further. Results of studies by Murray and Hitchcock (1969) and by Bryden (1971), in which recall strategies were studied, also show that the dichotomy of echoic vs. active short-term memory does not provide comparable results across different tasks and instructions.

METHOD

Subjects

Thirty-two female Ohio State University undergraduates served as *S*s in fulfillment of a course requirement.

Apparatus and Stimulus Materials

The *S* sat in a dimly lit room. Two 7.5-w G. E. night-light bulbs, 23 cm. apart and 70 cm. in front of *S*, served as cues for message recall. A red light on the right signaled recall of digits presented to the right ear and a white light on the left signaled recall from the left ear. The messages were presented through an Ampex Fine Line (F-44) stereo tape deck and (GSC) stereo headphones. Messages were composed of lists of randomly ordered monosyllabic digits (1, 2, 3, 4, 5, 6, 8, 9) presented at the rate of 1.8 digits/sec to each ear.

Procedure

Each *S* participated in two 45-min. sessions on 2 consecutive days.

Session 1. The purpose of this session was to acquaint *S* with the experimental tasks. The session was divided into four blocks in the following order:

1. Binaural shadowing (5 trials): On each trial *S* was required to shadow a list of 16 randomly ordered digits presented to both ears simultaneously.

2. Binaural shadowing with recall (32 trials): The *S* shadowed 32 new lists, each ending with a letter. This letter was always one of the four acoustically nonconfusable letters: B, F, K, and R. As soon as *S* repeated the letter the two cuing lights were flashed simultaneously. The *S* was then required to recall immediately the three digits that preceded the letter in their order of presentation. In order to maintain the uncertainty associated with the letter position in the list, the 32 lists varied in

length (4, 5, 7, 8, 10, 13, or 14 items). Each of the eight list lengths was presented four times.

3. Dichotic selective condition (10 trials): On each trial two different lists of randomly ordered digits were presented simultaneously, one to each ear, at the rate of 1.8 pairs/sec. The first trial was used to equate for each *S* the subjective loudness level of the two lists. There followed 10 trials on which *S* was instructed to shadow one of the lists, and only that list terminated with a letter. As soon as *S* repeated the letter, one of the two cuing lights was flashed. If the right red light appeared, *S* was required to report the last three digits presented to the right ear. If the left white light appeared, *S* was to recall the last three digits presented to the left ear. One half of the *S*s were informed that on all trials the light spatially corresponding to their shadowed message would appear, and one half of the *S*s were informed that on all trials the light corresponding to the nonshadowed message would appear.

4. Dichotic divided condition (10 trials): This condition was similar to the dichotic selective condition with one important exception: The target letter was now situated as the last item in the digit lists presented to the nonshadowed channel. At the onset of the expected cuing light, *S* was required to recall both the letter presented to the nonshadowed ear and the last three digits presented to the cued (expected) channel.

Session 2. The data collected in this session were used for data analysis. Each *S* performed 10 trials of binaural shadowing. Following these trials half of the *S*s received a block of 32 trials under the selective-attention condition and then 32 trials under the divided-attention condition. The order of the two conditions was reversed for the remaining *S*s. The 32 trials on each block were made up of an equal number of the eight list lengths, randomly ordered within a block. The major difference in these two conditions between Day 1 and Day 2 was that on Day 2, prior to the first trial, *S* was informed that now the expected cuing light would appear most of the time (rather than all the time as on Day 1), but that on some of the trials the other, unexpected, light would appear. In either case, *S*'s task was to recall the digits presented to the ear corresponding to the light that was flashed. Unknown to *S*, the relative frequency of the two lights was 75% for the expected channel and 25% for the unexpected one. Immediate recall of the letter that ended the shadowed (in the selective condition) or the nonshadowed (divided condition) message was also required.

On both days *S* was instructed to try and recall as many of the three digits as possible without missing any of the digits in the shadowing task. Intertrial intervals were approximately 10 sec. within a block with 3 min. between blocks.

Design

The 32 *S*s were assigned to one of four groups counterbalanced for the ear to which the shadowed message was presented (right vs. left) and for the expected recall channel (right vs. left). This resulted in two groups for which the expected recall message was the shadowed one (SE) and two groups for which the expected recall message was the nonshadowed one (NSE). All other variables were within *S*s. The design in terms of number of trials in each condition (selective vs. divided) and task (digits and letter recall) is summarized in Table 1. In this table the entries in each cell represent the number of trials for one *S* (from each of the two types of groups) from which her probability score was computed. Digit recall is denoted by D and letter recall by L. The letters EC and $\overline{\text{EC}}$ refer to expected channel and unexpected channel, respectively. Expectancy does not apply to the letter recall since the channel to which the letters were presented was determined by the condition only.

TABLE 1

NUMBER OF RECALL TRIALS OF LETTERS (L) AND DIGITS (D) IN EXPECTED (EC) AND UNEXPECTED ($\overline{\text{EC}}$) CHANNEL FOR AN *S* IN EACH CONDITION AND RECALL CHANNEL ON DAY 2

Condition	Recall channel			
	SE group		NSE group	
	Shadowed	Non-shadowed	Shadowed	Non-shadowed
Selective	24 D(EC) 32 L	8 D($\overline{\text{EC}}$)	8 D($\overline{\text{EC}}$) 32 L	24 D(EC)
Divided	24 D(EC)	8 D($\overline{\text{EC}}$) 32 L	8 D($\overline{\text{EC}}$)	24 D(EC) 32 L

Note. For the SE *S*s the shadowed channel contains the expected recall message, while for NSE *S*s the nonshadowed channel contains the expected recall message.

RESULTS

Only recall data collected on Day 2 were analyzed. Shadowing performance was almost errorless for all *S*s in all conditions and groups and was not further analyzed.

Digit Recall

Probability of recall on each recall channel, condition, and serial position was computed for each *S*. Preliminary analysis showed that both main effects and the interaction of right vs. left ear of shadowed and expected recall channel did not affect recall significantly. The data were then pooled into a 2 × 2 × 2 × 3 analysis of variance which compared two groups of *S*s (SE and NSE), two conditions (selective and divided), and two recall channels

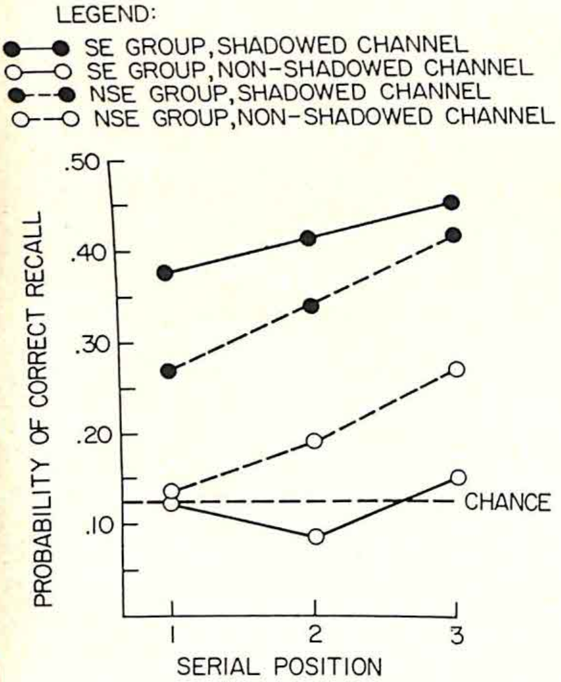


FIG. 1. Probability of correct recall as a function of group, recall channel, and serial order. (On the abscissa, 3 is the most recently presented digit.)

(shadowed and nonshadowed) across three serial positions. All main effects, with the exception of groups, F (1, 30) < 1.00, were significant beyond the .001 level.

Recall probability, when averaged across all serial positions, was .39 from the shadowed message, and this was superior to recall on the unshadowed message where the probability was .13, F (1, 30) $= 41.01$. In the selective condition, *S*s recalled each digit with probability of .32 compared to a much lower level of .21 in the divided condition, F (1, 30) $= 25.02$. Overall, recall probability of the most recently presented digit was greatest ($P = .32$) and recall probability of the earliest digit was least ($P = .23$), F (2, 60) $= 28.48$, $p < .001$.

TABLE 2

PROBABILITY OF DIGIT RECALL FOR THE TWO GROUPS IN THE TWO CHANNELS

Group	Recall channel		
	Shadowed	Nonshadowed	Average
SE	.414 (.396)	.119	.267
NSE	.342	.199 (.185)	.271

Note. For the SE *S*s the shadowed channel contains the expected recall message, while for NSE *S*s the nonshadowed channel contains the expected recall message. Numbers in parentheses are recall probabilities for the first eight recall trials from the expected channel.

The Groups × Recall Channel interaction was significant, F (1, 30) $= 4.94$, $p < .05$, and is described in Table 2. While for both groups digit recall is superior in the shadowed channel, the margin of superiority is greatly reduced for NSE *S*s. This interaction confirms the prediction that for a given recall channel, recall will be better for those *S*s who expect to recall from it. In this situation, recall from the shadowed channel was better for those *S*s expecting to be tested on it (SE *S*s), while recall from the nonshadowed channel was better for NSE *S*s. This relationship was practically identical for both the selective and divided conditions. Thus, although average recall accuracy was equivalent for the two groups (.53 for SE *S*s and .54 for NSE *S*s), performance at each channel was markedly different.

Performance of the two groups also differed over the three serial positions as indicated in Figure 1. The NSE *S*s, who were trained to anticipate digit recall in the nonshadowed channel, exhibited strong recency effects in recalling from either the shadowed or the nonshadowed channel. Digit recall of the SE *S*s, in contrast, did not change markedly with serial position in recalling digits from either channel. As a result the superiority of NSE *S*s to SE *S*s in recall from the nonshadowed channel was greatest for the most recently presented digit. Similarly, the fallibility of NSE *S*s in recalling earlier digits explains their relatively poor overall performance on the shadowed channel. This Group × Serial Position interaction was significant at the .01 level, F (2, 60) $= 5.95$, and did not change as a function of recall channel. The three-way interaction Group × Recall Channel × Serial Order was not significant, F (2, 60) < 1.00.

Also significant were the two two-way interactions, Conditions × Recall Channel, F (1, 30) $= 12.72$, $p < .01$, and Condition × Serial Order, F (2, 60) $= 8.38$, $p < .001$, as well as the three-way interaction, Condition × Recall Channel × Serial Order, F

$(2, 60) = 17.17$, $p < .001$. This last interaction is plotted in Figure 2, where the data show that when recalling digits from the nonshadowed channel the recall curves are low, with a slight recency effect, for both the selective- and divided-attention conditions. However the recall curve in the shadowed channel is markedly affected by the condition, with a large recency effect in the selective condition only. This result supports the assertion that expectancies (condition) can alter the nature of the basic processes.

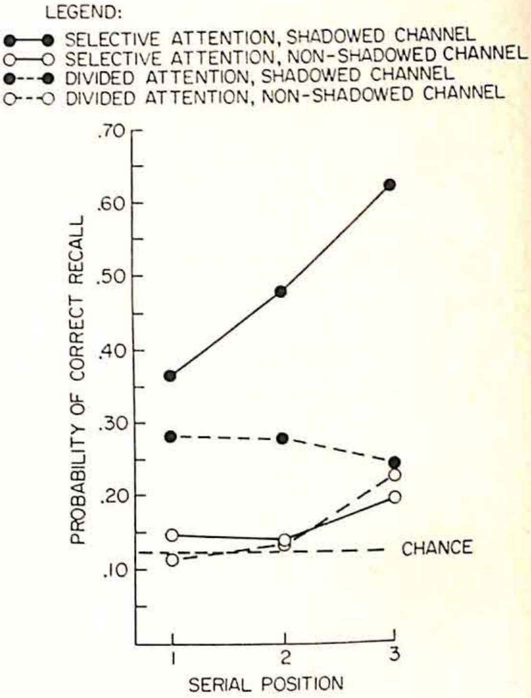


FIG. 2. Probability of correct recall as a function of condition, recall channel, and serial order. (On the abscissa, 3 is the most recently presented digit.)

Letter Recall

Again, preliminary analysis showed that both main effects and the interaction of right vs. left ear of shadowed and expected recall channel did not affect recall significantly. The data were then pooled into a 2×2 analysis of variance that compared the two groups of *S*s in the two conditions (for letter recall, recall channel coincided with condition, since expectancy did not apply to the letter recall). As predicted, neither the main effect of group, $F(1, 30) < 1.00$, nor the interaction, Group × Condition, $F(1, 30) < 1.00$, was significant. The predicted main effect of condition also failed to reach significance, $F(1, 30) = 1.95$, but the difference was in the predicted direction. Recall probability was slightly higher in the selective condition ($P = .832$), where the letter was the last item in the shadowed channel, than in the divided attention condition ($P = .763$), where the letter was presented to the nonshadowed channel.

DISCUSSION

The results of this study raise questions about the utility of passive filter-type models as valid descriptions of human information processing. Rather, they illustrate the importance of cognitive strategies based upon expectations or set in determining performance in a divided-attention task. The main result supporting this conclusion was the interaction between actual recall channel and the expected recall channel (i.e., group). This interaction reflects *S*'s ability to allocate more capacity to a nonshadowed channel when expecting to recall its content than when not. Similarly, the expectation to recall the shadowed message improves recall from that channel at the expense of the other channel. This result is contradictory to the hypothesis that the level of processing afforded each of the channels in dichotic listening is determined only by the operational task of shadowing one of the two channels. The finding is more consistent with that of Murray and Hitchcock (1969), namely, that shadowed-channel recall improves more if *S*s are informed in advance that they will be required to recall from that channel, than if they are uncertain as to the probed recall channel. However, in their study there was no complementary improvement for the nonshadowed message under prior cuing, as was found in this study. Their result may be due to the fact that in the latter condition articulatory shadowing was replaced by "writing down the contents" of the "shadowed" message.

It could be argued that the interaction between the actual and the expected recall channel is an artifact due to the smaller number of trials each *S* had to recall from the unexpected channel (8 vs. 24). To test for this possibility, recall probabilities were analyzed separately only for the first eight recall trials from the expected channel. The results from this analysis are presented in parentheses in Table 2. Clearly, the effect of the added practice causes little qualitative change in performance; indeed digit recall probability increases by an average of only .016.

Two additional aspects of the data suggest the role of instruction-induced expectancies. The first involves the near identical average digit recall levels of the two instructional groups across all conditions (main effect of groups nonsignificant). This average recall probability (Table 2) can be viewed as an index of capacity in this task, and the fact that it is not affected by the expectancy suggests that, given the proper set, *S*s utilize their limited attentional capacities to develop an optimal time-sharing strategy. Moray and Fitter (in press) reach similar conclusions on the basis of visual time-sharing performance.

The second line of evidence for the importance of expectancy comes indirectly from letter recall data. The message in which the letter appeared was determined by the condition only and was independent of the instructions governing expectancy (for both instructional groups the letter terminated the nonshadowed message in the divided condition and terminated the shadowed message in the selective condition). Therefore, unlike digit recall, which is affected by the instructions, letter recall should be equivalent for both groups. Letter recall probabilities for SE (.815) and for NSE (.780) supported this prediction.

Recall functions (Figure 1 and Figure 2) further suggest that both expectancy (group in Figure 1) and the location of the letter (selective and divided conditions in Figure 2) contribute to the determination of the "attentional strategy" adopted by *S*. The *S*s in the NSE group show a larger recency effect than *S*s in the SE group on both channels (Figure 1). If performance on the nonshadowed channel were dependent on the capacity of some passive sensory storage as Treisman and Riley (1969) claim, then the recall functions from this channel for the two groups should be identical or at least parallel. This is because expectancy in Treisman's (1964) model affects the recognition of specific items after they have entered an echoic storage of physical attributes (Swets & Kristofferson, 1970). These findings indicate that echoic storage capacity for random digit strings varies directly with the degree of attention allocated, as suggested by Neisser (1967) and Murray and Hitchcock (1969). The similar difference in recall from the shadowed channel suggests that attentional strategies with unconstrained materials affect storage after full identification (i.e., shadowing) has occurred. Therefore Norman's (1968) model might seem more appropriate since he posits his attention mechanism after STS. But neither Norman's model nor that of Deutsch and Deutsch (1963) accounts for recall differences *between* the two instructional groups, since expectancy here is in terms of input source rather than specific items that have prior memorial representations.

Figure 2 shows that the recency effect associated with short-term storage is manifested in recall from either channel only in the selective condition, i.e., when the letter is presented in the shadowed message. In contrast, when attention must be divided because the letter is presented to the nonshadowed channel, the recall from the nonshadowed channel is reflected by a flat function, suggesting a constant level of transfer of information into a long-term store. Thus, it can be argued, the position of the letter (selective or divided condition) determines the processing strategy used by *S* and thus affects his digit recall performance. When *S* must divide his attention in order to identify the letter in the nonshadowed channel the message encoded for shadowing purposes is transferred to long-term memory and discarded from STS where the limited capacity is shared with the nonshadowed message. In the selective condition *S*'s strategy is to store information in STS for as long as possible since the time-sharing requirements are smaller. This in turn results in a sharp recency effect with superior recall at all three serial positions. In both cases *S*'s strategy is the intervening variable that affects performance level. The idea that the storage of information for recall purposes can be independent of the level of processing is also supported by Bryden's (1971) results.

In other respects the present study replicates previous findings. In general, shadowed-channel information is recalled better than that of the nonshadowed one (Treisman, 1964); the serial position function for immediate recall of digits presented to the nonshadowed ear when the *S*s expect to recall from it (Figure 1) shows a distinct recency effect (Norman, 1969); the probability of letter recall from the nonshadowed channel approximates that found by Treisman and Riley (1969). Also, the lack of a primacy effect in all the recall functions is typical for a running memory span and has been replicated for dichotic presentations (Norman, 1969).

Taken together these results argue for a flexible strategy-dependent model of human information processing. This view implies that the body of data amassed by previous studies should be viewed as the manifestation of an attentional process in conjunction with a strategy. Unfortunately, in most previous studies it is hard to separate the two, and in those cases where the experimental task was changed between conditions (e.g., Norman, 1969), it is likely that different strategies underlie different performance functions. The results of this study are therefore viewed as more consistent with the active approach; they show that capacity of the active memory is limited (overall probability of correct recall being the same for both groups), but that allocation of attention can be directed by instructions as well as by classes of input.

While division of attention may be flexible, questions remain regarding the nature of the sampling process employed by *S*s. At least two possibilities have been offered: (*a*) continuous sampling of information from both channels with (Treisman, 1964) or without (Deutsch & Deutsch, 1963) alternation of one of the two input sources, and (*b*) discrete sampling of one channel at a time (e.g., Moray & Fitter, in press). For all practical implications the two types of sampling are indistinguishable when the switching time of a discrete sampling operation becomes infinitely small (Lindsay, Taylor, & Forbes, 1968). However, if the hypothesized sampling rate is slower, such as two switches per second (Broadbent, 1958), then it becomes meaningful to examine the conditional probabilities of recall of the last digit, given the letter is recalled when it appears in the other channel, P (D_3/L). Conditional and marginal probabilities for recall of the last digit when it is simultaneous with the last letter are presented in Table 3. In all cases the conditional probabilities are higher than the marginal probabilities. This then suggests that the allocation of attentional capacity is continuous or discrete with a high switching rate. The conditional probabilities also reflect the superiority of NSE *S*s in the nonshadowed channel, where they expected to be cued for recall. Thus, we are led to believe that allocation of capacity is continuous and affected by *S*'s expectancies.

TABLE 3

Conditional and Marginal Probabilities for Recall of Last Digit When Letter is Presented to Other Channel

Group	Selective condition: Nonshadowed channel		Divided condition: Shadowed channel	
	$P(D_3)$	$P(D_3/L)$	$P(D_3)$	$P(D_3/L)$
SE	.143	.225	.230	.266
NSE	.248	.401	.256	.300

REFERENCES

Broadbent, D. E. *Perception and communication.* New York: Pergamon, 1958.

Bryden, M. P. Attentional strategies and short-term memory in dichotic listening. *Cognitive Psychology*, 1971, **2**, 99–116.

Deutsch, J. A., & Deutsch, D. Attention: Some theoretical considerations. *Psychological Review*, 1963, **70**, 80–90.

Greenwald, A. G. Selective attention as a function of signal rate. *Journal of Experimental Psychology*, 1970, **86**, 48–52.

Harris, C. S., & Haber, R. N. Selective attention and coding in visual perception. *Journal of Experimental Psychology*, 1963, **65**, 328–333.

Hochberg, J. Attention, organization, and consciousness. In D. J. Mostofsky (Ed.), *Attention: Contemporary theory and analysis.* New York: Appleton-Century-Crofts, 1970.

Lewis, J. L. Semantic processing of unattended messages using dichotic listening. *Journal of Experimental Psychology*, 1970, **85**, 225–228.

Lindsay, P. H., Taylor, M. M., & Forbes, S. M. Attention and multidimensional discrimination. *Perception & Psychophysics*, 1968, **4**, 113–117.

Moray, N., & Fitter, M. A theory and the measurement of attention. In S. Kornblum (Ed.), *Attention and performance IV.* New York: Academic Press, in press.

Murray, D. J., & Hitchcock, C. H. Attention and storage in dichotic listening. *Journal of Experimental Psychology*, 1969, **81**, 164–169.

Neisser, U. *Cognitive psychology.* New York: Appleton-Century-Crofts, 1967.

Norman, D. A. Toward a theory of memory and attention. *Psychological Review*, 1968, **75**, 522–536.

Norman, D. A. Memory while shadowing. *Quarterly Journal of Experimental Psychology*, 1969, **21**, 85–93.

Swets, J. A., & Kristofferson, A. B. Attention. *Annual Review of Psychology*, 1970, **21**, 339–366.

Treisman, A. M. Selective attention in man. *British Medical Bulletin*, 1964, **20**, 12–16.

Treisman, A. M. Strategies and models of selective attention. *Psychological Review*, 1969, **76**, 282–299.

Treisman, A. M., & Geffen, G. Selective attention: Perception or response? *Quarterly Journal of Experimental Psychology*, 1967, **19**, 1–17.

Treisman, A. M., & Riley, J. G. A. Is selective attention selective perception or selective response: A further test. *Journal of Experimental Psychology*, 1969, **79**, 27–34.

(Received June 19, 1972)

Journal of Experimental Psychology
1973, Vol. 98, No. 2, 246–251

CHANGES IN VISUAL SPATIAL ORGANIZATION: RESPONSE FREQUENCY EQUALIZATION VERSUS ADAPTATION LEVEL[1]

WILLIAM STEINBERG AND ROBERT SEKULER[2]

Northwestern University

Two experiments tested alternative explanations for the apparent change in visual spatial organization produced by exposure to "adapting" or anchor stimuli. The alternative positions evaluated were that the change was perceptual (adaptation-level [AL] theory) or that it was the result of response bias (response-frequency-equalization theory). In Experiment I, judgments of spatial organization were obtained for 11 different matrices of dots in a series ranging from columnar organization to rowlike organization. The *S*s were exposed on alternate trials to dot patterns of extreme horizontal (rowlike) organization, to extreme vertical (columnar) organization, or to neither (control group). As predicted by both theories, judgments were systematically affected by the adapting stimuli. In Experiment II, the influence of response bias was nullified by a forced-choice procedure. Contrary to AL theory, once response bias was eliminated group differences vanished.

Adaptation is a change in sensory response which follows the presentation of some stimulus. The study of visual adaptation provides a powerful analytic tool for identifying the nature and character of sensory mechanisms which support various perceptual experiences (Weisstein, 1969). Adaptation-level (AL) theory provides one general framework which integrates, in a limited set of equations, many apparently unrelated phenomena connected with adaptation (Helson, 1964).

While adaptation may potentially follow exposure to any stimulus which differs from the current base-line level or pattern of stimulation (Andrews, 1964; Held, 1962), it is important to distinguish actual changes in sensory response from ersatz forms of adaptation. Only the former changes are likely to illuminate sensory processes. Sekuler and Erlebacher (1971) identified one possible artifact which could produce ersatz adaptation, the tendency of *S*s to use available responses with equal frequency. Among the experiments singled out as demonstrations of ersatz, rather than genuine, adaptation are several reported by Bell and Bevan (1968). In Bell and Bevan's first experiment, which may be considered as the prototype for all their experiments, *S*s judged the spatial organization of a series of dots arranged on the vertices of an imaginary matrix. When the vertical distances between dots are greater than the horizontal distances, the dots seem to be organized in rows. When the vertical distances between dots are less than the horizontal distances, the dots seem to be organized in columns. These observations demonstrate the operation of the gestalt principle of "proximity." Bell and Bevan used a graded series of dot matrices ranging from extreme rowlike organization through neutral organization (vertical distances equal to horizontal distances) to extreme columnlike organization. Each stimulus was judged, using the method of single stimuli, as rowlike or columnlike. The point of subjective equality (PSE) can be defined as the stimulus which appeared most nearly neutral (i.e., judged rowlike with $P = .5$). The PSE was determined for each of three groups. For the column group, every other stimulus to be judged was extremely columnar in its organization—such a stimulus was virtually certain to elicit a column judgment. For the row group, every other stimulus was extremely rowlike in its organization—being virtually certain to elicit row responses. A final group, the control group, was tested without the introduction of extreme anchor stimuli. The PSE of the column group was shifted away from the neutral stimulus of the series toward columnlike stimuli; the PSE of the row group was shifted in the opposite direction, toward rowlike stimuli. Bell and Bevan attributed the opposite shifts in PSE to sensory adaptation produced by repeated presentation of the extreme anchor stimuli. This finding, they claimed, was consistent with Helson's (1966) belief "that there are spatial adaptations, no less than chromatic adaptations, which are responsible for the way spatial patterns are formed [p. 173]." This claim has since been repeated by Avant and Bevan (1971) in their discussion of the Bell and Bevan experiments.

---

[1] Preparation of this article was aided by National Institute of Health Grants EY-00321 and NS-10094.

[2] Requests for reprints should be sent to Robert Sekuler, Cresap Laboratory of Neuroscience and Behavior, Northwestern University, Evanston, Illinois 60201.

Sekuler and Erlebacher (1971) proposed that the data of Bell and Bevan had nothing whatever to do with sensory adaptation, but merely resulted from response bias, the tendency for *S*s to use all available responses with approximately equal frequency. Response-frequency-equalization (RFE) theory argues that *S*s in Bell and Bevan's row group, having made a row response to every other (anchor) stimulus, would, whenever uncertain as to the correct response, make column responses. This would shift the PSE away from the neutral stimulus toward rowlike stimuli. The same argument could be applied to the column group, producing the obtained shift in the PSE toward columnlike stimuli. A more formal statement of this theoretical position can be found in Erlebacher and Sekuler (1971).

The present paper reports two experiments designed to discriminate between ersatz and genuine sensory adaptation. The first experiment is a replication of the original Bell and Bevan (1968) study, done to ensure the comparability of our stimuli and procedures to theirs. In the first experiment, like Bell and Bevan, we used the method of single stimuli. The second experiment, with the same stimuli as the first, uses a method drawn from the realm of objective psychophysics, the forced-choice method (Green & Swets, 1966). In the past this method has been useful for discriminating sensory effects from effects attributable to decision processes. In the present context, forced-choice procedures show that Bell and Bevan mistook the effects of response-frequency equalization for real sensory adaptation.

EXPERIMENT I

Method

Subjects. Thirty undergraduate students enrolled in an introductory psychology course at Northwestern University served as *S*s. Ten *S*s were randomly assigned to each of three groups; all *S*s in any group were run simultaneously.

Materials and procedure. The stimuli were eleven patterns of black, .5-in.-diam. dots. Each pattern contained 36 dots within an 8-in. square. Slides were projected onto a white beaded screen by means of a projector equipped with a tachistoscopic shutter. A black velvet mask covered most of the screen, leaving an irregular-shaped opening through which the slides were projected. On one side of the opening was a large photograph of the row anchor pattern labeled "rows"; on the other side was a similar photograph of the column anchor pattern, labeled "columns." The pattern on the photographs was also 8 in. square. The ratios of the horizontal to the vertical spacings of the dots for the 11 stimuli and the two anchors were identical to those used by Bell and Bevan (1968). These ratios in the columnar patterns were 1.11, 1.25, 1.44, 1.67, and 2.00, and in the row patterns were .90, .80, .70, .60, and .50. The neutral stimulus, of course, had a horizontal to vertical separation ratio of 1.00. Ratios for the column and row anchors were, respectively, 10.0 and .10.

The *S*s were seated along three concentric arcs centered on the screen. The average viewing distance was 15 ft., causing the stimuli to subtend, for the average *S*, approximately 2° 30′ of visual angle. The experiment was conducted in a partially darkened room. Light areas of the stimuli had a luminance of 60 cd/m^2 and had a contrast of approximately .80. Each slide was presented for 500 msec., followed by a 4.5-sec. interval during which *S*s could mark either "R" or "C" on their data sheets, indicating a judgment of row or column organization.

The three groups of *S*s were treated as Bell and Bevan (1968) treated their groups. The *S*s in all three groups received 55 experimental trials each, in five independent block randomizations of the 11 test stimuli. In addition, on every other trial the column group was given an extreme columnlike stimulus (described previously). The row anchor group received an extreme rowlike stimulus on the corresponding trials. The control group received no anchor stimuli, giving judgments on just the 55 experimental trials.

Results and Discussion

The proportion of row judgments for each *S* and stimulus was computed.[3] For purposes of data analysis, these proportions were transformed to twice the arc sine of the proportion's square root. When the proportion itself was either 1.0 or 0.0, the transformation was performed on values of .9 or .1, respectively. These substituted proportions are $\pm 1.0/2n$, where n is the number of observations upon which the proportion is based. Only data from 55 experimental trials have been analyzed since they are the only ones common to all groups. It might be mentioned, however, that row responses were made to virtually 100% of the interpolated anchor stimuli given the row group, and column responses were made to virtually all of the interpolated anchors given the column group.

The analysis of variance showed that the differences among groups, $F\ (2, 27) = 9.01$, and the main effect of stimulus type, $F\ (10, 270) = 80.07$, were both highly significant ($p < .01$). The interaction between stimulus type and group proved nonsignificant, $F\ (20, 270) = 1.56$, $.06 < p < .07$. As in the Bell and Bevan (1968) experiment, which the present study attempts to replicate, the proportion of row judgments is greatest for the column group, least for the row group, and intermediate for the control group.

Least squares fits to the z-transformed proportions were obtained in order to estimate the points of subjective equality (PSE) for each group. The equations of the best-fitting lines were: $z = -1.82s + 1.93$, $z = -1.43s + 1.15$, and $z = -1.90s + 2.21$, for the control, row, and column groups, respectively, where z is the unit normal deviate of the proportion of row responses, and s is the ratio between horizontal and vertical distances in a stimulus pattern. The root mean square errors about the fitted lines are .35, .35, and .34, respectively. Solving the least squares for $z = 0.0$ gives PSEs of 1.06, .80, and 1.16 for control, row, and column groups. Bell and Bevan (1968) report 1.06, .87, and 1.19 as the corresponding PSE values. The correspondence between the values in our experiment and those of Bell and Bevan is striking.[4] Just as in the Bell and Bevan study, then, the PSE for the column group is shifted toward columnar stimuli, while the PSE of the row group is shifted in the opposite direction.

To facilitate visual comparison between our data and those of Bell and Bevan (1968), Figure 1 shows our data plotted as Bell and Bevan gave theirs.

---

[3] For the row group, one presentation of the most extreme row pattern (other than the row anchor pattern) was lost because of mechanical failure. Thus, this pattern was only presented four times to the row group. A similar error caused the column group to receive only four presentations of the neutral pattern. These losses have been compensated by using percentages as the basic data for the experiment rather than frequencies.

[4] Although the numerical near identity of our data and those of Bell and Bevan (1968) is encouraging, too much ought not be made of that identity, since one cannot tell from their article how psychometric functions were fitted and PSEs estimated. William Bevan was kind enough to send us the data from the experiment by Bell and Bevan and has informed us that they fit their psychometric functions "with the aid of French curves." To insure comparability between our PSE estimates and the estimates for the Bell and Bevan experiment, we have obtained least squares fits to their data using the same procedure as for our own data. The PSEs which result from the least squares equations are 1.13, .82, and 1.23, for the control, row, and column groups in the Bell and Bevan experiment.

An examination of the residuals (predicted z values minus obtained z values) from our least squares fits indicates that there is a distinct nonlinear component in our data which remains after the z transformation (see Draper & Smith, 1966). This means of course that the original, untransformed data were not terribly well described by the integral of the normal curve. We believe that at least a large portion of this nonlinearity in z for the z-transformed data can be accounted for by substantial numbers of stimuli, at either end of the continuum, that lie essentially at the basement and ceiling of the psychometric function. This supposition is supported by the outcome of fitting linear functions to the z-transformed proportions of only the middle five stimuli, i.e., from 1.25 to .80, horizontal to vertical ratios. The best-fitting equations for these truncated least squares fits are: $z = -3.82s + 3.80$, $z = -2.32s + 1.81$, and $z = -3.57s + 3.83$, for control, row, and column groups, respectively, with average root mean square errors of .14. The PSEs derived from these truncated psychometric functions are .99, .78, and 1.07, preserving quite nicely the magnitude and order of differences among the three groups while shifting all the PSEs somewhat toward lower horizontal to vertical ratios.

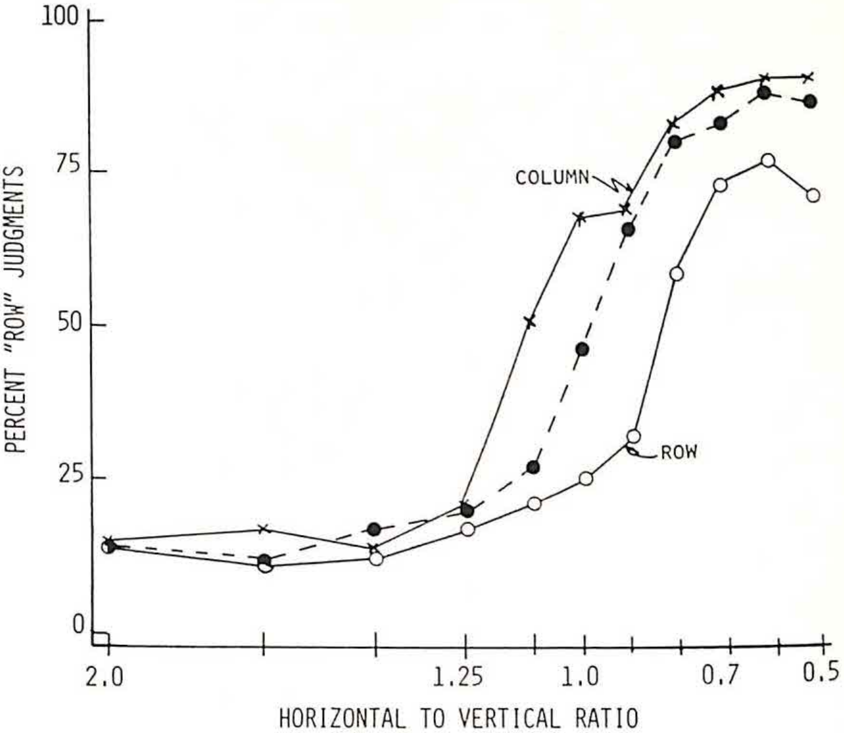


FIG. 1. The percentage of row responses for each group as a function of the ratio of horizontal to vertical separation of dots in the stimuli. (Crosses show data for the column group, filled circles for the control group, and unfilled circles for the row group.)

EXPERIMENT II

The first experiment established the comparability of our stimuli and apparatus to those used by Bell and Bevan (1968). It will be recalled that both RFE theory and AL theory made the same predictions for that experiment. The second experiment uses the same stimuli and apparatus in a situation where adaptation-level theory and response-frequency-equalization theory make quite different predictions. On each trial, two dot patterns are presented and *S* makes a forced-choice response, indicating which of the two patterns is the more nearly neutral (i.e., least like either rows or columns in its organization). For all three groups in this experiment, the comparisons of interest are between a physically neutral dot pattern and one of the eleven patterns used in the first experiment, including the neutral pattern itself. For one group, row group, on every other trial *S*s are given two identical, extreme row anchor patterns, again being asked to judge which of the two is more neutral. The repeated presentation of extreme row stimuli should, according to adaptation-level theory, shift the adaptation level toward rowlike stimuli. This would cause physically neutral stimuli, as a result, to appear slightly columnar. The experimental consequences of this shift in AL may be represented as follows: When a slightly rowlike pattern and a neutral pattern are presented on the same trial, the slightly rowlike pattern, being closer to the AL, should be judged—erroneously—as the more neutral of the two. In general, AL theory predicts a shift in the PSE for the row group away from the physically neutral stimulus toward a more rowlike pattern. For another group in Experiment II, the column group, *S*s see, on alternate trials, two identical extreme columnar patterns. Again, the repeated presentation of extreme column stimuli should, if AL theory is correct, shift the AL toward columnar stimuli, causing a physically neutral pattern to appear rowlike. By reasoning similar to that given above for the case of the row group, for the column group, adaptation-level theory predicts a shift in the PSE away from a physically neutral pattern toward one that is columnar. In Experiment II, a third, control, group receives no extreme anchor or adapting stimuli. For all groups, on the trials when a neutral stimulus is compared against one of the eleven test patterns, the temporal position, first or second, of that neutral stimulus was randomized. According to response-frequency-equalization theory, this randomization should spread the effects of any possible response bias (i.e., bias produced by using too many first or second judgments) evenly across all neutral and nonneutral stimuli alike. With response bias thus cancelled out, if there are differences among the groups—similar to those of the first experiment—we may conclude that the data of the first experiment and those of Bell and Bevan (1968) were the product of genuine sensory adaptation. However, with response bias cancelled out, should the differences among the groups disappear, we must conclude that the differences in the first experiment and those by Bell and Bevan were the result of ersatz adaptation, of the tendency of *S*s to use available responses with equal frequency.

Method

Subjects. Thirty-three undergraduate students from lower-level psychology courses served as *S*s. They were paid $2 for their service and, in addition, were informed about their eligibility for a $5 bonus that would be given to the single *S* whose judgments of neutrality had the fewest errors. The instructions about the bonus, which was actually paid, served to enhance attention and concern for accuracy in judgment. Eleven *S*s were assigned randomly to each group.

Materials and procedure. With the following exceptions, the materials and procedure were the same as those described for Experiment I. The *S*s in all groups were given the following sequence of events on each trial: One slide was presented for .5 sec.; then, after an interval of 2 sec., another slide was presented for .5 sec. The *S*s had the following 9 sec. to mark either "1" or "2" on their data sheets to indicate whether they thought the first or second pattern presented was the more neutral. The *S*s in the control group received 55 trials, representing 5 independent block randomizations of the 11 test stimuli, each being paired with the neutral stimulus. The *S*s in the row group received the same order of these 55 trials and, in addition, on every other trial were presented with two extreme row patterns, identical to the row anchors used in Experiment I. The *S*s in the column group were treated as those in the row group except that they were shown pairs of extreme column stimuli on alternate trials rather than extreme row stimuli. On those trials, for all groups, where a neutral stimulus was being compared to one of the 11 test stimuli, the temporal order of the neutral stimulus, first or second presented on that trial, was randomized. The *S*s were run one group at a time with a seating arrangement similar to that of Experiment I.

Results

Only the data from the 55 trials per *S* that all groups had in common were analyzed. Judgments of the extreme anchor stimuli were not analyzed since they were included only to produce a possible differential shift in AL. The basic data were the number of times each of the 11 test stimuli was judged less neutral than the physically neutral stimulus. The frequencies were converted to proportions and then transformed to twice the arc sine of the proportion's square root. These transformed scores were entered into an analysis of variance. As in Experiment I, proportions near 1.0 or 0.0 were adjusted to .9 or .1. The analysis of variance showed only a significant main effect of stimulus type, $F\ (10, 300) = 41.85$, $p < .01$, with neither the main effect of group nor the interaction between groups and stimulus type approaching statistical significance, $F\ (2, 30) = .96$, $p > .50$ and $F\ (20, 300) = 1.11$, $p > .30$. The data, expressed as percentages, are shown in Figure 2. The same curve has been drawn through the data for all groups. This curve represents the mean values for all three groups. The figure is presented this way to permit the reader to determine for himself, independently of the analysis, whether there are systematic differences among the groups. This visual presentation plus the analysis of variance leads us to conclude that the large and powerful group difference found in Experiment I and in the Bell and Bevan (1968) experiment has evaporated under the forced-choice procedure used in the present experiment.

GENERAL DISCUSSION

The article by Bell and Bevan (1968) reported four different experiments. Each experiment demonstrated the effects of extreme

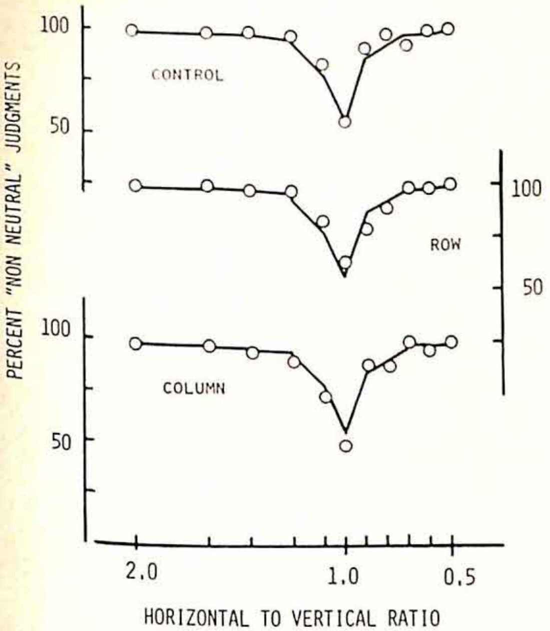


FIG. 2. Proportion of nonneutral responses as a function of the ratio of horizontal to vertical separation of dots in the stimuli. (Data for the control group are shown in the top section, data for the row group in the middle section, and for the column group in the bottom section of the figure. Each set of data has been displaced by a constant amount [.75] above its lower neighbor to facilitate inspection. The curve shown in every panel is the grand mean.)

anchor stimuli upon a different gestalt principle of perceptual organization. The experiment that we have replicated was one which was designed to show that the contribution of "proximity" could be altered by adaptation level. Other experiments purported to show similar effects for the gestalt principles of "similarity," "good continuation," and "closure." The present study shows that the effects reported by Bell and Bevan (1968) and alleged to show changes in the perceptual operation of proximity are merely the product of response bias, as suggested some time ago by Sekuler and Erlebacher (1971). This leads us to propose that most, if not all, of the experiments reported by Bell and Bevan are similarly explicable in terms of response-frequency equalization. Thus, rather than supporting Helson's (1966) proposition that there exist forms of spatial adaptation analogous to chromatic adaptation, we believe Bell and Bevan's data support the claim that a wide variety of psychophysical experiments may be "contaminated by the artifactual effects of *S*'s tendency to use response with approximately equal frequency [Sekuler & Erlebacher, 1971, p. 311]."

It must be made clear that we are not taking the position that all evidence of spatial adaptation is to be written off as being the result of response bias. We believe that there are several valid demonstrations of spatial adaptation, some measured under forced-choice conditions. These experiments make rival hypotheses extremely difficult to conjure up. We also believe that wherever possible, forced-choice procedures ought to be used in studies of visual adaptation or, at an absolute minimum, precautions should be taken to remove response-frequency equalization as a rival hypothesis for explaining findings that presumably are meant to illuminate working of sensory processes. This caveat must be seriously considered regardless of the psychophysical procedure used—whether it be method of constant stimuli, limits, or adjustment. The present experiments do show that, given converging operations, it is possible to distinguish sensory effects from those which are due to response processes or biases.

REFERENCES

ANDREWS, D. P. Error-correcting perceptual mechanisms. *Quarterly Journal of Experimental Psychology*, 1964, **16**, 104–115.

AVANT, L. L., & BEVAN, W. Adaptation level and visual space perception. In M. H. Appley (Ed.), *Adaptation-level theory: A symposium.* New York: Academic Press, 1971.

BELL, R. A., & BEVAN, W. The influence of anchors upon the operation of certain gestalt organizing principles. *Journal of Experimental Psychology*, 1968, **78**, 670–678.

DRAPER, N. R., & SMITH, H. *Applied regression analysis.* New York: Wiley, 1966.

ERLEBACHER, A., & SEKULER, R. Response frequency equalization: A bias model for psychophysics. *Perception & Psychophysics*, 1971, **9**, 315–320.

GREEN, D. M., & SWETS, J. A. *Signal detection and psychophysics.* New York: Wiley, 1966.

HELD, R. M. Adaptation to rearrangement and visual-spatial aftereffects. *Psychologische Beitrage*, 1962, **6**, 439–450.

HELSON, H. *Adaptation level theory.* New York: Harper & Row, 1964.

HELSON, H. Some problems in motivation from the point of view of the theory of adaptation level. *Nebraska Symposium on Motivation*, 1966, **14**, 137–182.

SEKULER, R., & ERLEBACHER, A. The invalidity of 'Invalid Results From the Method of Constant Stimuli': A common artifact in the methods of psychophysics. *Perception & Psychophysics*, 1971, **9**, 309–311.

WEISSTEIN, N. What the frog's eye tells the human brain. *Psychological Bulletin*, 1969, **72**, 157–176.

(Received June 19, 1972)

Journal of Experimental Psychology
1973, Vol. 98, No. 2, 252–259

EFFECT OF IRRELEVANT SURROUNDING MATERIAL ON SPEED OF *SAME–DIFFERENT* JUDGMENT OF TWO ADJACENT LETTERS[1]

LESTER E. KRUEGER[2]

City College of the City University of New York

To test whether *same–different* judgments depend on a fast global comparison process for *same* decisions and a serial analytic process for *different* decisions, the two letters presented simultaneously for comparison were bracketed with irrelevant material in a symmetrical or asymmetrical fashion, i.e., different material was placed next to each letter. The symmetrical material affected *same* and *different* response times (RT) equally, whereas the asymmetrical material, which was intended to give each letter a different framework and thus prevent global comparison, did increase RT more for *same* than for *different* judgments. More false *different* than false *same* responses were made, especially when an asymmetrical surround was included, which suggests an explanation of the typically longer latencies for *different* than for *same* responses: Imperfectly encoded noisy features produce spurious mismatches, so *S* must double check and verify all but the largest perceived differences before responding *different*.

When *S* judges whether two items are the *same* (e.g., AA) or *different* (e.g., AB), his response time (RT) typically is shorter for *same* than for *different* stimulus pairs (e.g., Nickerson, 1965). This finding has led some investigators to postulate that, to a degree, different stimulus-comparison processes subserve *same* and *different* decisions. According to the dual-process model, *same* judgments are based mainly on image- or template-matching and involve a global holistic comparison of the two items in a pair, whereas *different* judgments are based mainly on a slower serial process (Bamber, 1969) in which *S* may name the two letters (Krueger, 1970) and may recheck the pair in order to determine exactly where the difference lies (Tversky, 1969). In most *same–different* studies, the two items in a pair are presented successively, but when two uppercase letters are presented simultaneously, shorter RT for *same* than for *different* pairs has been found by Krueger (1973) and Egeth and Blecker (1971), though not by Posner and Mitchell (1967).

In the present study, to test the dual-process model the two letters presented simultaneously for comparison were each undergirded with the same irrelevant material (symmetrical condition), with different material (asymmetrical condition), or with no material at all (no-surround condition). It was predicted that the irrelevant material would increase RT in general, but *same* RT would increase more than *different* RT in the asymmetrical case, and *different* RT more than *same* RT in the symmetrical case. The global comparison process postulated by the dual-process model to underlie *same* decisions ought to be disrupted in the asymmetrical condition, with *S* forced to rely more on the slower serial identification process for *same* as well as *different* judgments. In the symmetrical condition, on the other hand, the global comparisons of *same* pairs ought to proceed largely unabated, whereas the naming and identification involved in *different* pairs ought to be slowed much more. Krueger (1970) presented vertical lines symmetrically to both the left and right of each of two adjacent letters and found that the height of the lines affected *dif-*

---

[1] This study was supported by National Institutes of Health Grant HD-04869-02. The author is grateful to Jeffrey Wildfogel and Harry Rosenberg for collecting and tabulating the data, to Richard C. Atkinson and Ira T. Kaplan for helpful discussions, and to Harold L. Hawkins and Michael I. Posner for incisive comments on an earlier draft of this report.

[2] Requests for reprints should be sent to Lester E. Krueger, Department of Psychology, City College of the City University of New York, New York, New York 10031.

ferent but not *same* RT. *Different* RT was longest when the vertical lines had the same height as the letters, thus producing minimal figure-ground separation between lines and letters. The results suggested that *S*s identified or named the two letters before making a *different* response, but relied on a more global comparison encompassing both the letters and the symmetrically arranged surrounding lines in the *same* case.

METHOD

In Experiments I and II, an underline was placed under the left or the right letter in a pair in the asymmetrical condition and under both letters in the symmetrical condition (see Figure 1a). Asymmetrical stimuli were randomly intermixed with symmetrical and no-surround stimuli in Experiment I, but formed a separate block, presented either before or after the symmetrical and no-surround trials, in Experiment II. In Experiment III, asymmetrical trials again formed a separate block, but the underline was replaced by + and = which were placed farther away from the letters than the underline (see Figure 1b). Further, to eliminate any visual imbalance that might induce eye movements, the + and = symbols, which had been chosen because of their similar visual size and "weight," were placed both above and below each of the two letters in both the symmetrical and asymmetrical conditions. Symmetry in Experiment III thus did not involve the presence or absence of a symbol, but was confined to the feature aspect: In an asymmetrical stimulus, there was a + above and below one letter and = above and below the other letter.

Apparatus. Stimuli were presented on a Gerbrands tachistoscope (Model T-3B-1), which has six timers and three exposure fields (all set to maximum intensity of 4 mL.), each with a semiautomatic card changer. Information on RT (measured in milliseconds by a specially built electronic timer) and which response *S* made was recorded on punched paper tape for subsequent computer analysis. To start each trial, *S* pressed a foot pedal linked to a remote-start switch on the tachistoscope; the switch was rendered inoperative between trials so that *S* could not initiate a trial before being permitted to do so by *E*.

Stimulus materials. Each stimulus consisted of two capital letters, along with irrelevant material (underline, +, or =), typed in 12-pt. IBM Courier 72 and centered on a 4 × 6 white index card (see Figure 1). Five blank spaces separated the two letters. At 79 cm. from *S*, the two letters subtended a horizontal visual angle of 1.25°. The vertical visual angle for the letters was .2°. In Experiment III, each letter with a surround was separated by .5° from the + or = immediately above and below, and the total vertical visual angle was 1.5°.

Two sets of 400 letter pairs were put on cards. In one set, used in Experiments I and II, on 100 pairs each, (*a*) the left letter alone was underlined; (*b*) the right letter alone was underlined; (*c*) both letters were underlined; and (*d*) neither letter was underlined. In the second set, used in Experiment III, similarly on 100 pairs each, (*a*) + appeared above and below the left letter and = above and below the right letter; (*b*) + appeared above and below the right letter and = above and below the left letter; (*c*) the same symbol (+ on 50 pairs, = on 50 pairs) appeared above and below both letters; and (*d*) no symbol appeared above or below either letter. Thus, both sets of cards contained 200 asymmetrical pairs, 100 symmetrical pairs, and 100 no-surround pairs. Further, in both sets, half of the cards contained two identical letters and half contained two different letters. The letters were randomly sampled from the 26 letters of the alphabet.

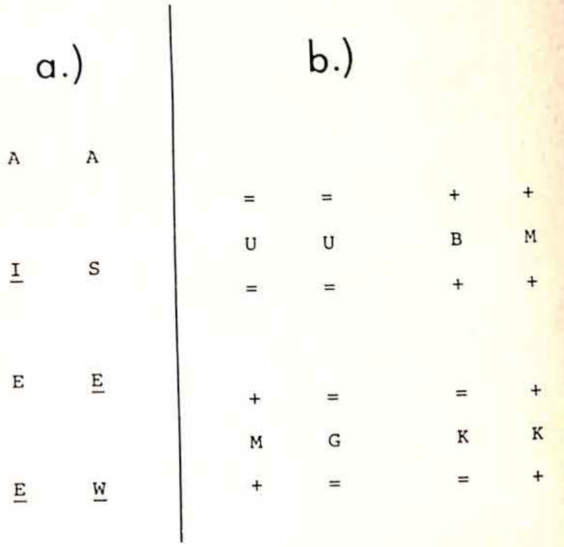


FIG. 1. (a) The four types of stimuli used in Experiments I and II (one no surround, one symmetrical, and two asymmetrical); (b) four of the five types of stimuli used in Experiment III (two symmetrical and two asymmetrical). (The no-surround type, also used in Experiment III, is not shown in Panel b, but may be seen in Panel a. Except for photographic reduction, the examples are exact replicas of stimuli shown to *S*s.)

The 400 stimulus cards used in each experiment were divided into 10 sets of 40 cards each. In each set, 20 *same* and 20 *different* pairs were intermixed randomly, except that no more than four cards in succession could be either *same* or *different*. In Experiment I, each set of 40 cards contained 10 each of left underlined, right underlined, both underlined, and neither underlined stimuli. The 10 sets of 40 cards were presented in a different random order for each *S*, but the random order within each set did not vary across *S*s. In Experiment II, one block of five sets of 40 cards contained only asymmetrical stimuli (left underlined, right underlined), and the other five sets contained symmetrical (both underlined) or no-surround stimuli. For half of the *Ss*, the five asymmetrical sets were presented before the other five sets, and after the other five sets for the other half of the *Ss*. A corresponding arrangement of five asymmetrical sets and five symmetrical and no-surround sets was used in Experiment III.

Procedure. The *Ss* were instructed to keep their eyes fixated midway between the two letters in a pair, as indicated by a prior, small black fixation dot. On each trial, the fixation point appeared for .8 sec., then the letter pair appeared for .6 sec., after which the fixation point reappeared for 2.5 sec. The *S*'s response did not terminate the stimulus display, which always remained in view for .6 sec. The *E* changed the stimulus card after *S* responded, so that 3–4 sec. typically elapsed between the end of one trial and the beginning of the next. A buzzer sounded briefly when *S* pressed the incorrect button or pressed before the stimulus card appeared. In Experiment I, *S* received eight practice trials before seeing the 400 stimulus cards. In Experiments II and III, *S* received four practice trials before the first five sets of trials and four practice trials before the second five sets of trials. All RTs which exceeded 3 sec. or on which *S* pressed the incorrect button were excluded before RTs were averaged.

Half of the *Ss* pressed the right button for *same* and the left button for *different*, and half pressed the left button for *same* and the right button for *different*. The *Ss* were asked to respond as quickly as possible, but not at the expense of accuracy. Each *S* was paid $1.50 for the 50-min. session, plus a bonus of $.50 if his error rate was below 5% and his speed of response was in the top quartile for all *Ss*.

Subjects. Ten City College of New York undergraduates served as *Ss* in each experiment. All had 20/30 vision or better as tested with a Snellen chart.

RESULTS

Mean RT was significantly higher for symmetrical than for no-surround trials for all three experiments combined (645 vs. 638 msec.), F (1, 27) = 5.26, $p < .05$; in Experiment III analyzed separately (601 vs. 584 msec.), F (1, 9) = 9.02, $p < .025$; but not in Experiments I and II. The error rate was higher, but not significantly so, for symmetrical than for no-surround trials for all three experiments combined (4.5% vs. 3.6%), F (1, 27) = 3.92, $p < .10$, and there was no significant effect in any experiment analyzed separately. For neither error rate nor RT was the Symmetrical vs. No-Surround × *Same–Different* interaction significant in any analysis. On error rate, the symmetrical condition was virtually the same as the no-surround condition on *different* judgments (3.4% vs. 3.3%), but was slightly higher on *same* judgments (5.5% vs. 3.9%). On RT, the increase for the symmetrical condition as compared with the no-surround condition was virtually the same for *different* as for *same* decisions (8 vs. 7 msec.).

Because the symmetrical and no-surround conditions produced such similar results and because, as compared with the asymmetrical condition, both were predicted to produce a similar RT profile (faster *same* than *different* decisions), they were combined in all analyses involving the asymmetrical condition. The term "symmetry" thus will refer to the comparison of the asymmetrical condition with the combined symmetrical and no-surround conditions.

As Figure 2 shows, mean RT was slightly higher for the asymmetrical condition in each experiment, but symmetry did not significantly affect RT for the three experiments combined nor in any individual experiment. The Symmetry × *Same–Different* interaction, however, was significant in Experiment I, F (1, 9) = 64.15, $p <$ .001; Experiment III, F (1, 9) = 9.95, $p <$ .025; and for the three experiments combined, F (1, 27) = 34.90, $p < .001$. The interaction stems largely from the tendency for the increase in RT for asymmetrical (vs. symmetrical) trials to occur on *same* rather than on *different* pairs. For *same* stimuli analyzed separately, the asym-

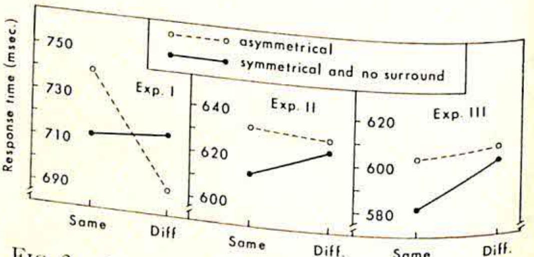


FIG. 2. Mean response time, in milliseconds, for *same* and *different* stimuli, for asymmetrical and the combined no-surround and symmetrical conditions, in Experiments I, II, and III. (Underlines were used in Experiments I and II, and + and = in Experiment III. The asymmetrical trials were intermixed with other trials in Experiment I, but were put in a separate block of trials in Experiments II and III. $N = 10$ each, for Experiments I, II, and III.)

TABLE 1

Mean Percentage Error Rate, by Stimulus Condition, in Experiments I, II, and III

Experiment	Stimulus condition			
	Symmetrical and no surround		Asymmetrical	
	Same	*Different*	*Same*	*Different*
I	4.8	3.0	6.6	2.6
II	5.4	3.3	6.7	3.8
III	3.8	3.7	5.0	2.6
I, II, and III	4.7	3.3	6.1	3.0

Note. The errors listed under *Same* represent false *different* responses. The errors under *Different* represent false *same* responses. $N = 10$ each, for Experiments I, II, and III.

metrical stimuli showed significantly longer mean RT than the symmetrical and no-surround stimuli, both for Experiment I, $F\ (1, 9) = 9.10$, $p < .025$, and for all three experiments combined, $F\ (1, 27) = 7.67$, $p < .01$. For *different* stimuli analyzed separately, the asymmetrical RT was significantly shorter in Experiment I, $F\ (1, 9) = 6.79$, $p < .05$, but not for all three experiments combined ($F < 1$).

Same and *different* RT did not differ significantly in any experiment or for all three experiments combined. Significant differences were found, however, on separate analyses for asymmetrical and for symmetrical and no-surround conditions. For the asymmetrical condition, *different* RT was significantly shorter than *same* RT by 16 msec., $F\ (1, 27) = 5.18$, $p < .05$, for all three experiments combined; and by 49 msec., $F\ (1, 9) = 7.84$, $p < .025$, in Experiment I. For the combined symmetrical and no-surround conditions, *same* was faster than *different* RT by 12 msec., $F\ (1, 27) = 5.26$, $p < .05$, for all three experiments combined; and by 12 msec. in Experiment II, $F\ (1, 9) = 5.12$, $p < .05$; and 20 msec. in Experiment III, $F\ (1, 9) = 8.75$, $p < .025$.

As Figure 2 shows, the effect of symmetry on *same* vs. *different* RT is much more apparent in Experiment I, where asymmetrical stimulus cards were intermixed with symmetrical and no-surround ones, than in Experiments II and III, where the asymmetrical stimulus cards formed a separate homogeneous block. The three-way interaction, Experiments I, II, III × Symmetry × *Same–Different*, was significant, $F\ (2, 27) = 6.51$, $p < .025$. Separate analyses between each pair of experiments revealed that the relationship between symmetry and *same–different* was significantly different between Experiments I and II, $F\ (1, 18) = 7.85$, $p < .025$; and between Experiments I and III, $F\ (1, 18) = 17.89$, $p < .001$; but not between Experiments II and III ($F < 1$). The three experiments did not differ significantly on overall mean RT.

On error rate, shown in Table 1, symmetry had no significant effect, but *same* vs. *different* did. The *S*s made significantly more errors on *same*, i.e., false *different* responses, than *different* letter pairs for all three experiments combined, $F\ (1, 27) = 21.60$, $p < .001$; in Experiment I, $F\ (1, 9) = 11.16$, $p < .01$; and in Experiment II, $F\ (1, 9) = 8.33$, $p < .025$. The Symmetry × *Same–Different* interaction was significant for all three experiments combined, $F\ (1, 27) = 5.38$, $p < .05$, but not for any individual experiment. The interaction reflects the fact that *same* and *different* differed more in the asymmetrical condition. The triple interaction, Experiments I, II, III × Symmetry × *Same–Different*, was not significant ($F < 1$). The three experiments did not differ significantly on overall error rate.

Discussion

It was predicted, based on the dual-process model, that (*a*) adding symmetrical irrelevant material would increase RT more for *different* than for *same* judgments, whereas (*b*) adding asymmetrical material would increase RT more for *same* than for *different* judgments. The first prediction, concerning symmetrical material, was not confirmed. Krueger (1970) found that placing two vertical lines of letter height next to each letter, increased RT more for *different* than for *same* responses; in the present experiments, the added material (an underline or the more distant + or =) may simply have been less effective in obscuring letter features. In the symmetrical and no-surround conditions, however, the typical finding of shorter *same* than *different* RT was confirmed. The shorter RT for *same* decisions cannot be attributed simply to a greater readiness or preference for responding *same* because even in the symmetrical and no-surround conditions, where such a response bias for *same* ought to be strongest, *S*s more often responded *different* (50.7%) than *same* (49.3%), making more false *different* than false *same* responses.

The second prediction was confirmed: Adding asymmetrical material did increase RT more for *same* than for *different* judgments. It was predicted, more specifically, that the asymmetrical material would merely eliminate the typical difference in RT between *same* and *different* by forcing *S*s to rely on a slower serial identification process for *same* as well as for *different* judgments. The prediction was largely upheld in Experiments II and III. The deviation from the prediction in Experiment I, where in the asymmetrical condition *different* RT was significantly shorter than *same* RT, may reflect a tendency for *S*s to incorporate the underline with the material being compared.

In Experiments II and III, the asymmetrical stimuli were grouped in a separate block of trials so that *S*s could more consistently adopt a rule to filter out the asymmetrical material. Further, the underline used in Experiments I and II may have been difficult to filter out; lying quite close to the letters, it may have formed an integral unit, or gestalt, with the letters (cf. Garner & Felfoldy, 1970). Some *S*s reported that the underline disrupted their comparisons of pairs such as OQ and IJ. Therefore, in Experiment III, the irrelevant material was placed further away from the letters and was made more visually balanced, so that it might be more easily filtered out.

The changes made in Experiments II and III were effective in rendering RT approximately equal for *same* and *different* in the asymmetrical condition, as Figure 2 shows. On error rate, however, the asymmetrical material did accentuate the difference between *same* and *different* stimuli in all three experiments. False *same* responses, representing errors on *different* pairs, occurred at about the same rate for asymmetrical and for symmetrical and no-surround stimuli, but false *different* responses were much higher for asymmetrical than symmetrical or no-surround stimuli. In the more difficult asymmetrical condition, *S*s may simply have felt a greater overall readiness to respond *different*, but the preponderance of *different* over *same* responses was only slightly greater in the asymmetrical condition (51.6% vs. 48.4%) than in the symmetrical and no-surround conditions (50.7% vs. 49.3%). Rather, it may be that *S*s did not succeed in completely filtering out the irrelevant asymmetrical material even in Experiments II and III. The irrelevant material in Experiment III (+ and =), even if nonintegral with the letters, could have interfered with *same* pairs in the asymmetrical condition by providing an alternative, competing stimulus which signaled a *different* response (cf. Hawkins & Shigley, 1972). Egeth (1967) noted that it may be particularly difficult to ignore irrelevant dimensions that engender competing responses.

The notion of competing responses or conflicting cues, when applied to the set of relevant dimensions, suggests one possible explanation of the longer RT for *different* responses. Since only one feature need differ in order for a stimulus pair to be labeled *different*, there may be many features that typically are identical in a *different* pair and thus constitute cues for the incompatible *same* response. No such incompatible cues can impede a *same* response because all features must be identical for a stimulus pair to be labeled *same*. If incompatible *same* cues interfered with the *different* response, however, *S*s ought to have been less accurate as well as slower on *different* than on *same* pairs, yet *S*s made more errors on *same* than on *different* pairs.

The higher error rate for *same* pairs also would seem to contradict the notion that *S*s responded faster to *same* pairs because it is easier to form representations of two identical letters than of two different letters. In choice RT tasks, immediately repeated stimuli typically are responded to faster than nonrepeated stimuli, which suggests that RT may be lower for *same* than for *different* judgments because the encoding of one item in a pair produces a short-term facilitation that allows the second item to be read in more rapidly when it matches the first (cf. Eichelman, 1970; Entus & Bindra, 1970). Such facilitation ought to lead to more accurate as well as more rapid comparison of identical or *same* stimuli, but the error rate was higher for *same* than for *different* letter pairs.

The effect of the asymmetrical material on error rate may yet be revealing as to why *same* judgments typically are faster than *different* judgments. First, it should be noted that presenting two capital letters simultaneously for comparison led to a significantly higher error rate for *same* than for *different* pairs not only in the present study, but also in previous studies by the author (Krueger, 1970, Experiment I; Krueger, 1973). The fact that adding irrelevant asymmetrical material, which introduced actual physical differences into the letter-pair situation, accentuated the preponderance of false *different* responses, suggests that even in the normal no-surround case, perceived differences often arose when the two letters being compared were actually identical.

The spurious differences, in turn, suggest that feature information reaching the comparison process was faulty or that the comparison process itself was error prone. Nickerson (1965) made no allowance for such "noise" in the comparison process when he remarked that the shorter RT for *same* than for *different* judgments is puzzling because *S* presumably need only discover one differing aspect in order to know that two letters are different, "whereas to determine that they are identical he must examine *every* detail to satisfy himself that there is *no* way in which they are not the same [p. 17]." The seeming paradox will be dispelled by showing how an unreliable comparison process might present greater difficulties for *different* than for *same* judgments.

The comparison process may be unreliable mainly because *S* does not always wait until the representation of one stimulus is perfectly clear and noise free before matching it with that of the other stimulus. Posner and Mitchell (1967) presented two letters simultaneously and found *same* judgments to be about 70 msec. faster for physical matches (e.g., AA) than for name matches (e.g., Aa). Beller (1970) proposed that the physical matches resulted from preattentive processing, whereas the name matches resulted from focal attentive processing. Neisser (1967) distinguished preattentive processes from focal attention by noting that "the first is fast, crude, wholistic, and parallel while the second is deliberate, attentive, detailed, and sequential [p. 10]." The preattentive processes segregate the figural units whose detailed features are clarified and fleshed in by focal attention.

A model of the same–different *comparison process.* A model will now be presented to explain the typically shorter latencies for *same* judgments. In the present model, it is postulated that the comparison process begins early, at the preattentive stage, and takes the form of a running tally of feature matches and mismatches. Mismatches involving large, quite discriminable differences will receive more weight in the tally than those involving small differences. It further is postulated that *S* responds as soon as the tally exceeds a criterion level. The criterion level is held to depend on the expected or judged quality of the stimulus representation. A higher criterion is needed for comparisons at the preattentive stage where the features are crude and inaccurate, thereby producing spurious matches and mismatches.

A crucial postulate in the model is that the presumably random processes, or noise, that degrade the stimulus at the preattentive stage will very rarely convert a physical mismatch into a perceived match; there simply are too many different values that a typical feature may take for a random process to have much chance of making two different values identical. Much more frequently, the noise will change a physical match into a perceived mismatch.

It is postulated, therefore, that *S* will accept all feature matches as real because of the very low probability that they are due to chance fluctuations, but will be much more cautious on feature mismatches. The *S* will respond *same* as soon as all features are represented in the tally and all indicate matches. The *S* will respond *different*, however, not simply upon registering a single feature mismatch, even though a single difference qualifies a stimulus pair as *different*. Rather, *S* will wait until the number and magnitude of the mismatches pushes the tally past the value attainable by noise alone. The criterion is successively relaxed as the visual image becomes sharpened during preattentive processing and focal attention, so that a *different* pair containing slight, but discriminable differences will eventually meet the criterion.

If the criterion is relaxed too rapidly, spurious differences may exceed the criterion, thereby prompting a false *different* response. False *same* responses ought to occur very rarely because of the rarity of spurious matches. This analysis would account for the greater number of false *different* than false *same* responses in the present study; the false *same* responses that did occur might be ascribed largely to the occasional confusion that *S*s experience as to which button to press, rather than to spurious matches.

Nickerson (1969) presented a related model. A counter cumulates difference information until either a count criterion is met, thereby prompting a *different* response, or a time criterion is met, thereby prompting a *same* response. The basic model pertains to the single dimension case, but Nickerson does propose an extension to the multiattribute case in which the count rate would depend both on the number of differing dimensions and on the magnitude of the difference on each. Nickerson's counter cumulates only difference information, however, whereas in the present model the tally encompasses feature matches as well as differences. Further, Nickerson (1969) minimized the role of noise: "even zero differences between stimuli will produce some count, although the mean count rate in this case will be relatively low [p. 259]." In the present model, on the other hand, the noise factor plays a prominent role; it is because of noisy, unprocessed features that the difference count does not start small and grow larger, as in Nickerson's case, but rather, when differences are not present or are difficult to detect, starts large and grows smaller as further processing clarifies and sharpens the visual image. Finally, Nickerson's model pertains only to situations where RT is longer for *same* than for *different* responses, whereas the present model predicts shorter RT for *same* in some situations and shorter RT for *different* in other situations.

The present model predicts shorter RT for *same* than for *different* responses when the physical differences are difficult to detect, and longer RT for *same* than for *different* when the differences are easy to detect. The uppercase letters used in the present study provide an example of a difficult discrimination, because the letters in a *different* pair typically are nearly identical in size and amount of shading and may even have many identical features (e.g., EF, OQ). With difficult-to-detect differences, which would fall within the range of differences that can be produced by internally generated noise alone, all *different* pairs would require processing beyond the preattentive stage in order for the response criterion to be met. Many *same* pairs also would require further processing, so as to dispel spurious mismatches, but at least some *same* pairs would be processed clearly enough to yield a response at the preattentive stage. Thus, the response criterion generally would be met earlier on *same* than on *different* pairs.

With easy-to-detect differences, *same* RT ought to be reduced somewhat because some minor mismatches might now be tolerated in the tally, but the *same* response would nevertheless require an exhaustive interrogation of features. *Different* RT would decrease much more in the easy-discrimination condition because the difference between two stimuli now would be too large to attribute to chance, and *S* would self-terminate processing as soon as a feature mismatch was registered. A *different* response thus would be triggered very early in processing, usually before matches could be registered on all features, as required for a *same* response.

Several investigators have indeed found shorter RT for *same* than for *different* judgments when the differences are difficult to detect, and shorter RT for *different* than for *same* when the differences are easy to detect. Bindra, Donderi, and Nishisato (1968) found shorter RT for *same* when the lines being compared varied minimally in length, but shorter RT for *different* when the differences in length were larger and thus easier to discriminate. The same pattern of results held regardless of whether the two lines in a pair were presented simultaneously or successively. Nickerson (1967) found *same* decisions to be much faster than *different* decisions when the difference involved only one attribute (size, color, or shape), but not when the difference involved two or three attributes, in which case *different* decisions were faster in some conditions. Snodgrass (1972) found shorter RT for *same* than for *different* judgments of complex matrix patterns, whether presented simultaneously or successively. When simple patterns were presented, RT became shorter for *different* than for *same* responses in the simultaneous condition, though not in the successive condition.

Conclusion. The present article provides a more detailed rationale for the view that RT typically is shorter for *same* than for *different* because *same* judgments are based mainly on global holistic comparisons, whereas *different* judgments are based mainly on slower serial processes (cf. Bamber, 1969; Krueger, 1970; Tversky, 1969). In the present model, the global comparisons are attributed to Neisser's (1967) first-stage preattentive processes and the serial processes to the second-stage focal attention. There is an element of irony in the present model, in that *different* judgments are held to be typically slower than *same* judgments, not because differences are more difficult to see, but rather because they are so easy to see, even when they do not exist, thus forcing *S* to double check all but the largest perceived differences before responding *different*.

REFERENCES

BAMBER, D. Reaction times and error rates for "same"–"different" judgments of multidimensional stimuli. *Perception & Psychophysics*, 1969, 6, 169–174.

BELLER, H. K. Parallel and serial stages in matching. *Journal of Experimental Psychology*, 1970, 84, 213–219.

BINDRA, D., DONDERI, D. C., & NISHISATO, S. Decision latencies of "same" and "different" judgments. *Perception & Psychophysics*, 1968, **3**, 121–130.

EGETH, H. Selective attention. *Psychological Bulletin*, 1967, **67**, 41–57.

EGETH, H., & BLECKER, D. Differential effects of familiarity on judgments of sameness and difference. *Perception & Psychophysics*, 1971, **9**, 321–326.

EICHELMAN, W. H. Stimulus and response repetition effects for naming letters at two response–stimulus intervals. *Perception & Psychophysics*, 1970, **7**, 94–96.

ENTUS, A., & BINDRA, D. Common features of the "repetition" and "same–different" effects in reaction time experiments. *Perception & Psychophysics*, 1970, **7**, 143–148.

GARNER, W. R., & FELFOLDY, G. L. Integrality of stimulus dimensions in various types of information processing. *Cognitive Psychology*, 1970, **1**, 225–241.

HAWKINS, H. L., & SHIGLEY, R. H. Irrelevant information and processing mode in speeded discrimination. *Journal of Experimental Psychology*, 1972, **96**, 389–395.

KRUEGER, L. E. Effect of bracketing lines on speed of "same"–"different" judgment of two adjacent letters. *Journal of Experimental Psychology*, 1970, **84**, 324–330.

KRUEGER, L. E. Effect of stimulus frequency on speed of "same"–"different" judgments. In S. Kornblum (Ed.), *Attention and performance, IV*. New York: Academic Press, 1973.

NEISSER, U. *Cognitive psychology*. New York: Appleton-Century-Crofts, 1967.

NICKERSON, R. S. Response times for "same"–"different" judgments. *Perceptual and Motor Skills*, 1965, **20**, 15–18.

NICKERSON, R. S. "Same"–"different" response times with multi-attribute stimulus differences. *Perceptual and Motor Skills*, 1967, **24**, 543–554.

NICKERSON, R. S. "Same"–"different" response times: A model and a preliminary test. *Acta Psychologica*, 1969, **30**, 257–275.

POSNER, M. I., & MITCHELL, R. F. Chronometric analysis of classification. *Psychological Review*, 1967, **74**, 392–409.

SNODGRASS, J. G. Matching patterns vs. matching digits: The effect of memory dependence and complexity on "same"–"different" reaction times. *Perception & Psychophysics*, 1972, **11**, 341–349.

TVERSKY, B. Pictorial and verbal encoding in a short-term memory task. *Perception & Psychophysics*, 1969, **6**, 225–233.

(Received June 22, 1972)

Journal of Experimental Psychology
1973, Vol. 98, No. 2, 260–263

STORAGE OF EVENTS AND EVENT FREQUENCIES: A COMPARISON OF TWO PARADIGMS IN MEMORY[1]

WILLIAM C. HOWELL[2]

Rice University

Comparable word lists were presented aurally to four groups of *S*s at each of four list lengths. Groups were distinguished in terms of retention task (free recall vs. frequency estimation) and instructional set (also recall vs. frequency). Words occurred at frequencies of 0, 1, 2, 4, 6, and 10 in each study list, and *S* responded to each word as it occurred by writing it out. Retention of frequency information turned out to be independent of set, while free recall did depend upon set to a significant degree. List length was of far greater importance in free recall than in frequency estimation, but word frequency was an important factor in both.

Event frequency is an important variable in at least two broad areas of memory research. One deals with *S*'s ability to retain the information conveyed by a particular event class (a word, a nonsense syllable, etc.); greater frequency presumably yields better retention (Underwood, 1969a). The other focuses on the concept of frequency per se; greater frequency leads *S* to perceive an event class as more frequent, probable, or likely (Hintzman, 1969, 1970; Howell, 1970; Underwood, Zimmerman & Freund, 1971). Both of these outcomes are well established and both could reflect a single underlying process—trace strength or multiplexing, for example (Hintzman & Block, 1971). Several studies, however, suggest that somewhat different processes might be involved (Underwood, 1969b; Howell, 1970).

Previous research leaves unanswered the question of whether memory enhancement and frequency perception are mediated by the same process. One principal reason is that "memory" and "frequency-perception" paradigms have not typically been employed in the same experiment. In the memory paradigm, *S* is instructed to remember *events* and, in fact, is tested for this at retrieval. In various adaptations of the frequency-perception paradigm he is tested using a *frequency* estimation or discrimination task after having been set for event memory (e.g., Hintzman, 1969), frequency storage (e.g., Erlick, 1961, 1962, 1963), or some vaguely described memory test (Underwood, Zimmerman, & Freund, 1971). Underwood (1969b) did compare event memory and frequency responses following similar lists of stimuli, but *S*s performed the two tasks successively and in each case under the appropriate set. The results showed high, but not complete agreement between the two measures, leading Underwood to suggest that "trace strength" and the "frequency attribute" are not identical.

The purpose of the present study, then, was to compare the two paradigms directly, in an effort to determine whether it is indeed necessary to postulate more than one frequency process. Different groups of *S*s were set for either event memory or frequency perception, observed the same word lists, and were tested on either a memory or frequency task. If the two paradigms simply involve a different read out of the same basic quantity (e.g., trace strength), instructions should have no appreciable effect on subsequent retrieval.

Frequency instructions pose one potential difficulty. If *S* knows he will be tested on this attribute, and if the task is otherwise not too demanding, he may actually maintain a running *count* of events as they occur.

---

[1] This research was supported by National Science Foundation Grant GB 25584. The assistance of Deborah L. Vandell and Martha M. Van Parys in collecting and analyzing the data is gratefully acknowledged.

[2] Requests for reprints should be sent to William C. Howell, Department of Psychology, Rice University, Houston, Texas 77001.

The intuitive frequency process of interest here would thus be masked by the highly learned counting skill. Since the point at which *S* would abandon an overt counting strategy is not readily determinable, the variable of *memory load* was introduced orthogonally to the other factors: Lists included 10–80 different words at comparable prescribed frequencies. This variable was not of interest in its own right since it is obviously confounded with such other factors as event spacing and list length. It was introduced only to assure that difficulty levels were included above the point of overt counting and below the point at which *S*s might simply "give up."

METHOD

Materials and apparatus. For each of the memory load conditions (10, 25, 50, and 80 events, respectively), four word samples were assembled by random selection from the 1–10 per million categories of the Thorndike-Lorge tables.[3] Word length was limited to 3–10 letters. Within each sample, six frequencies (0, 1, 2, 4, 6, and 10) were assigned at random to an equal number of words. Word lists were then constructed by a random ordering of all the items thus assembled, the only limitation being that no word could appear more than twice in succession in any list. The result was sixteen lists of 46, 115, 230, or 368 items apiece, four at each memory load condition. Thus at Load 10, for example, a list contained 2 words at each of the five nonzero frequencies; at Load 80, it contained 16 words at each of these same frequencies.

All word lists were recorded verbally on magnetic tape for auditory presentation at a rate of approximately 12/min. As each word was played back through the loudspeaker, *S* recorded its occurrence in pencil on a piece of paper exposed through a small window. The paper was moved beneath this aperture after each response so as to remove from view all previous responses. This same masking arrangement was used during the test phase to record frequency and free-recall responses individually.

Subjects and design. A total of 92 male and 100 female college students were recruited and paid for their participation (one session of less than 1-hr. duration). Assignment to groups was random with the restriction that, insofar as possible, sex was equated over groups. Sixteen groups were defined by the factorial combination of memory load (10, 25, 50, and 80 items), set (*memory* vs. *frequency*), and response (*recall* vs. *frequency estimation*). Thus there were 12 *S*s per group, three of whom performed on the same list. Word frequency, of course, was a within-*S* variable with five levels. For purposes of experimental control, the design was viewed as a single 5 × 4 × 2 × 2 factorial experiment with one within-*S* and three between-*S* treatments. For analytic purposes, however, it was treated as two 5 × 4 × 2 factorial experiments, since the response variable represented two very different dependent measures. The key issue was whether congruent sets produce superior performance on the retrieval tasks, and if so, whether the frequency functions are comparable. If a single memory process is involved, set should have no differential effect in either case; if more than one process is involved, one or both congruent sets (memory–memory [MM] or frequency–frequency, [FF]) should yield better retrieval than its incongruent control, FM or MF. If frequency controls different processes in the two paradigms, there should be some difference in obtained frequency functions.

Procedure. The *S*s served individually or in isolated pairs. Each was seated at a desk, was read the instructions appropriate for his condition, and was presented with each of the words in his list in a single familiarization trial. During this trial he heard each word only once, and was shown concurrently that same word printed on an index card. Instructions noted that these were the only words he would hear during the subsequent presentation, that each could occur any number of times, and that he would ultimately be required to recall them (memory set) or estimate their frequency (frequency set). He was further instructed to write each word, within 5 sec. of its occurrence, on the blank paper appearing in the window of the device on his desk.

After the familiarization trial, he followed the prescribed procedure throughout one presentation of the entire experimental list. For lists in excess of 46 items, he was given a 1-min. rest period after every 46 words. Following the experimental presentation, a set of instructions appropriate for either free recall or frequency estimation was read, and *S* was given 15 min. to complete the retrieval task. For free recall, *S* was told simply to list all the words he could remember. For frequency estimation he was instructed to write before each word presented in the window the number of times he thought it had occurred, including zero, in the previous list. Each of the words in the list, together with an appropriate number of zero-frequency items, was then presented individually.

At the conclusion of the retrieval task, each *S* was asked whether he could describe any particular strategy that he used to aid his memory. The nature of the experiment was then fully explained and his cooperation was solicited in not informing other potential *S*s of critical details.

RESULTS AND DISCUSSION

Recall. Mean percentages of recall scores are summarized in Table 1. Over all load

---

[3] Actually, 12, 30, 60, and 96 were drawn in order to provide some words that could occur with zero frequency.

TABLE 1

Mean Percentage of Recall under All 40 Memory Conditions

Group	Word frequency	Memory load				
		10	25	50	80	*M*
Frequency–Memory	1	58	25	31	17	25
	2	79	53	48	28	41
	4	71	72	58	38	51
	6	92	65	61	51	58
	10	100	85	79	54	69
	M	80	60	55	38	
Memory–Memory	1	71	35	38	20	31
	2	83	50	59	35	47
	4	83	80	58	53	61
	6	92	92	68	65	72
	10	100	82	79	61	72
	M	86	68	61	47	

levels and virtually all frequency levels, the congruent MM set yielded consistently superior performance at retrieval. An analysis of variance supported this conclusion, $F\ (1, 88) = 5.98, p < .025$ (one-tailed test). Load level and frequency were, as expected, also highly significant factors, $F\ (3, 88) = 38.18$, $p < .01$, and $F\ (4,\ 352) = 86.36$, $p < .01$, respectively. None of the interactions approached significance.

These results confirm the usual finding that recall is sensitive to event frequency. Considering the methodological differences, the functions are not inconsistent with those reported earlier by Underwood (1969b). More important, however, is the fact that congruent instructions enhance recall over and above the frequency effect but to a fairly comparable extent over all frequencies. This would suggest that *S* relies heavily, but not exclusively, on the representation of frequency (strength or multiple traces) in the recall paradigm.

The residual advantage of congruent instructions probably reflects an additional process, such as a scheme for organizing stimuli, that *S* employs under a memory set. Additional evidence favoring such an interpretation is provided by correlations computed between order of recall and word frequency for MM and FM groups. If frequency representation were the only quantity governing recall, one would expect events to be read out roughly in order of decreasing frequency. Such was indeed the case for the *S*s set for frequency (FM): ρ was computed as .40, which is significantly greater than zero, $t = 2.23, p < .05$. In contrast the ρ for the superior MM groups was $-.18$, which does not even approach significance, $p > .20$.

Frequency estimation. Mean frequency judgments are summarized in Table 2. All groups were able to do a respectable job of judging frequency, even under the highest memory load conditions. Mean estimates tended to undervalue high and overvalue low frequencies, as has often been reported in other contexts (e.g., Peterson & Beach, 1967).

Clearly, direct retrieval of frequency information was less dependent on set than was recall. Only at frequencies of 6 and 10 did the congruent FF *S*s show even a slight superiority over those set erroneously for a recall task. Also in contrast to the recall data, memory load had a negligible effect on performance. Apparently the frequency representation builds up with remarkable consistency, irrespective of whatever other strategies *S* may be led to adopt. These conclusions were substantiated by the analysis of variance: The only variable having a significant effect on frequency estimates was word *frequency*, $F\ (4,\ 352) = 289.53$, $p < .01$. In contrast the values

TABLE 2

Mean Frequency Estimates under All 48 Frequency Conditions

Group	Word frequency	Memory load				
		10	25	50	80	*M*
Memory–Frequency	0	0.00	.07	.15	.15	.09
	1	2.08	2.72	1.98	2.28	2.24
	2	2.88	3.15	2.99	3.47	3.24
	4	5.04	5.73	4.50	4.81	4.87
	6	6.04	6.32	5.37	5.91	5.81
	10	7.92	9.08	7.18	6.90	7.38
	M	3.99	4.51	3.70	3.90	
Frequency–Frequency	0	0.00	.15	.08	.02	.06
	1	2.00	1.60	1.99	2.54	2.03
	2	2.33	2.63	3.23	3.58	3.25
	4	3.71	4.13	4.61	5.56	4.94
	6	5.50	4.92	6.20	7.14	6.42
	10	9.21	7.43	8.50	8.64	8.42
	M	3.79	3.48	4.10	3.58	

for *set*, F (1, 88) = .02, and *load*, F (3, 88) = .65, did not even approach significance.

Any direct comparison between the frequency functions obtained under the two tasks is necessarily speculative since different measures are involved. Moreover, recall is limited at the upper end, while frequency judgment is not. Below this ceiling, however, there is little to suggest that different processes underlie the two obtained functions. Certainly both *could* arise from a common process.

The main conclusion to be drawn from the composite of the above data, then, is that some representation of event frequency—presumably trace strength or multiple traces—builds up as events are repeated, irrespective of specific memory or frequency instructions. From this, an *S* can read out frequency directly, subject only to whatever limitations there are on his ability to transform such information into numerical responses. He may also use this representation in retrieval if forced to do so. However, there are probably more efficient retrieval schemes to which he can (and apparently will) turn if provided with an appropriate initial set. In any case, instructions stressing frequency produce consistently inferior recall when compared to memory instructions, whereas both sets produce essentially the same frequency responses.

One further result is perhaps worthy of some comment. The memory load variable, confounded though it was, had a substantial effect on recall and none whatsoever on frequency estimation. This would suggest that the *frequency* representation of an event, however coded, is relatively undisturbed either by other events in storage or by their recovery; retrieval of the attributes defining an *event*, on the other hand, is highly susceptible to interference. Another conclusion to be drawn from the failure of memory load to affect frequency estimation is that even the smallest load was sufficiently demanding to obviate overt counting. Had a counting strategy been used, performance at Load 10 would have been considerably more accurate than at higher levels.

REFERENCES

Erlick, D. E. Judgments of the relative frequency of a sequential series of two events. *Journal of Experimental Psychology*, 1961, **62**, 105-112.

Erlick, D. E. Perception of the most frequent category of a random series as a function of the number of categories. *Journal of Experimental Psychology*, 1962, **63**, 115-118.

Erlick, D. E. Effects of method of displaying categories on the perception of relative frequency. *Journal of Experimental Psychology*, 1963, **66**, 316–318.

Hintzman, D. L. Apparent frequency as a function of frequency and the spacing of repetitions. *Journal of Experimental Psychology*, 1969, **80**, 139–145.

Hintzman, D. L. Effects of repetition and exposure duration on memory. *Journal of Experimental Psychology*, 1970, **83**, 435–444.

Hintzman, D. L., & Block, R. A. Repetition and memory: Evidence for a multiple-trace hypothesis. *Journal of Experimental Psychology*, 1971, **88**, 297–306.

Howell, W. C. Intuitive "counting" and "tagging" in memory. *Journal of Experimental Psychology*, 1970, **85**, 210-215.

Peterson, C. R., & Beach, L. R. Man as an intuitive statistician. *Psychological Review*, 1967, **68**, 29–46.

Underwood, B. J. Attributes of memory. *Psychological Review*, 1969, **76**, 559–573. (a)

Underwood, B. J. Some correlates of item repetition in free-recall learning. *Journal of Verbal Learning and Verbal Behavior*, 1969, **8**, 83–94. (b)

Underwood, B. J., Zimmerman, J., & Freund, J. S. Retention of frequency information with observations on recognition and recall. *Journal of Experimental Psychology*, 1971, **87**, 149–162.

(Received June 26, 1972)

Journal of Experimental Psychology
1973, Vol. 98, No. 2, 264–270

EFFECTS OF STIMULUS NOVELTY AND DIMENSIONAL SALIENCY IN HUMAN SHIFT LEARNING[1]

NICHOLAS J. ESPOSITO[2]

State University of New York College at Cortland

In 1968 Slamecka argued that shift-learning paradigms which employed geometric designs and human *S*s might have had confounded variables which could have been responsible for shift-learning results. While Slamecka's criticisms are logical possibilities, he did not provide direct experimental evidence for two of them: stimulus novelty and primary stimulus generalization. The present article argued against Slamecka's primary stimulus generalization criticism and experimentally tested the effects of stimulus novelty using the partial change and total change paradigms. Results showed significant main effects for saliency and type of shift and significant interactions between novelty and type of shift and saliency and type of shift. These results were interpreted in terms of the 1963 Zeaman and House attentional theory. Slamecka's criticism regarding stimulus novelty was partially justified.

Slamecka (1968) reviewed the various discrimination learning paradigms involving transfer of training in human beings and concluded that the basic design and all of its variations allowed explanations other than those based upon mediational processes. Two variations are of particular interest here: the *partial change design* and the *total change design*. In studies employing the partial change design, if *S* sorts cards with respect to forms in the preshift task (say circles vs. squares), then for the postshift task these cues of the form dimension are changed (e.g., to hexagons vs. triangles) while all other dimensions and their cues are unchanged. In the postshift task, the *S* must learn either an intradimensional (ID) shift (e.g., learning to sort hexagons vs. triangles, the changed and therefore novel cues of the form dimension) or an extradimensional (ED) shift (e.g., learning to sort the unchanged cues of a different dimension such as color).

In the total change design, all cues of all the dimensions are changed from the preshift task to the postshift task while the dimensions remain the same. For example, if form and color were the dimensions employed, the cues in the preshift task might be circles vs. squares and green vs. red. The cues in the postshift task might be hexagons vs. triangles and blue vs. yellow. Again, in the postshift task *S* must learn either an ID shift or an ED shift. For either type of shift, the relevant dimension in the postshift task would have new cues.

The results of studies using the partial change design (i.e., that ID shifts are learned in fewer trials than ED shifts with adult *S*s) may be due to stimulus novelty rather than mediational or attentional processes. The partial change design may favor ID shifts over ED shifts because the cues which are changed in the postshift task (and which are therefore novel) are on the dimension which was relevant during the preshift task. If the novelty of the changed cues attracts *S*'s attention, then *S*s should be more likely to respond to these novel cues. Concomitantly, *S*s should be less likely to respond to the unchanged (nonnovel) cues of the other dimensions. In order to execute an ID shift, an *S* is required, in the postshift task, to sort with regard to novel cues which are on the same dimension that was relevant in the preshift task. In order to execute an ED shift, *S* is required to sort with regard to some dimension other than the one that was relevant in the preshift task, i.e., a new relevant dimension, but one with unchanged cues. Therefore, type of shift is confounded with novelty in the partial change design.

---

[1] This research was conducted while the author held a Faculty Research Fellowship from the Research Foundation of State University of New York.

[2] Requests for reprints should be sent to Nicholas J. Esposito, Department of Psychology, State University of New York College at Cortland, Cortland, New York 13045.

In the total change design, since all cues of all dimensions are changed in the postshift task, cues of one dimension are no more novel than those of any other dimension. However, while the total change design controls for stimulus novelty, Slamecka (1968) claimed that it does not control for possible primary stimulus generalization effects when employed with geometric forms. Slamecka claimed that generalization effects are important because they allow a nonmediational interpretation of results of transfer of training studies even if the total change design were used. In particular, Slamecka stated that a primary stimulus generalization response might be viewed as a single-stage S–R counterpart of a dimensional mediator. Defining primary stimulus generalization "as an intrinsic tendency to respond to all stimuli lying on the same physical continuum as the original training stimulus [p. 436]," Slamecka conjectured that this tendency might channel *S*'s responses during the postshift task toward the dimension which was relevant in the preshift task. Therefore, an ID shift would be facilitated since the proposed channeling of *S*'s responses is consistent with the demands of an ID shift, and an ED shift would be hindered since the channeling of *S*'s responses is inconsistent with the demands of an ED shift.

Slamecka's (1969) ultimate solution to the novelty and generalization problems was to employ a total change design using words rather than geometric forms as stimuli. The total change design precluded any novelty effects and the words precluded any primary stimulus generalization effects. Under these conditions, he found fewer trials to criterion in the ID than in the ED shift condition using adult *S*s.

Though the criticisms which Slamecka (1968) has made regarding novelty and primary stimulus generalization are theoretically and logically possible and are based upon implications from experimental results, and though the total change design using words as stimuli does seem to preclude any confounding due to novelty and primary stimulus generalization, a closer analysis of these criticisms might be useful for the following reasons. First, the total change design with words cannot be used with preverbal children or animals, whereas the total change design with geometric forms can be so used. Second, there has been no direct experimental test of the novelty criticism. The novelty criticism was based upon implications from experimental results such as those of Isaacs and Duncan (1962) which showed an ID shift to be learned faster than a reversal shift, both of which were learned faster than an ED shift. While these results might be expected as a consequence of stimulus novelty, they might also be explained by the fact that the reversal shift is in essence a negative-transfer paradigm while an ID shift is not. Therefore, the Isaacs and Duncan study is not a good direct experimental test of the novelty criticism. Third, since there has been no direct experimental test of the novelty criticism, one cannot be sure if novelty effects are in fact present when the partial change design is used. Therefore, one does not know if the results of past studies using the partial change design are vitiated because of confounding. Fourth, any effects that novelty might have on discrimination shift learning would be theoretically interesting.

The results of studies using the total change design would, presumably, have no alternative interpretations if there were no possibility of primary stimulus generalization leading to a single-stage, nonmediational, S–R interpretation. There is evidence that the existence of these primary stimulus generalization effects as defined by Slamecka do not have firm experimental support.

Slamecka's (1968) definition of primary stimulus generalization as an "intrinsic tendency to respond to all stimuli *on the same physical continuum as the original training stimulus* [emphasis added]" and his subsequent argument (both on p. 436) appear to assume that even in the single-stage S–R model it is necessary for *S*s to attend to dimensions *before* primary stimulus generalization can account for the results of discrimination shift learning studies. That is, the definition of primary stimulus generalization which Slamecka adopts appears to assume that if *S* is presented with a stimulus bearing cues of small, green, and circle, then *S* will only pay attention, for instance, to green when the stimulus is reinforced and will subsequently show generalization along the color dimension only. Jeffrey (1965); Kalish (1969); Tighe, Glick, and Cole (1971); Graf and Tighe (1971); and Tighe and Frey (1972) presented evidence that animals and young children, who probably best fit the single-stage S–R model, would tend to view the above stimulus as a small-green-circle complex, i.e., a complex which is undifferentiated with regard to dimensions. Generalization, in this latter case, would depend upon the similarity of a new undifferentiated complex to the old complex. Thus, animals and young children do not seem to attend to a single dimension and then generalize along that dimension. Consequently, if Slamecka's definition is used, primary stimulus generalization seems not to be a single-stage counterpart of dimensional mediation as he suggested, but another name for dimensional mediation.

However, with the total change design, one must still consider primary stimulus generalization from cue to cue. For instance, suppose the relevant dimension in the preshift task were color, with the cues being red and green, and suppose that *S*s do attend to dimensions. Response A might be the correct response for red and Response B for green. Assume that the cues in the postshift task were yellow, which is more similar to red; and blue, which is more similar to green. Then, in the postshift task *S* would tend to give Response A to yellow and Response B to blue. If *S* were executing an ID shift and if yellow and green were assigned Responses A and B, respectively, then the ID shift would be facilitated. For an ED shift, since the dimensions are usually orthogonal in such tasks, both yellow and green would be paired with Response A half of the time and with Response B half of the time, so that stimulus generalization effects would not produce a facilitative effect.

One solution to the problem of ID shift facilitation is for half of the *S*s in the ID condition to have Response A assigned to yellow and Response B to blue so that there would be facilitation, and for the other half to have Response B assigned to yellow and Response A to blue so that there would be hindrance. It is hoped that the facilitation and hindrance would balance out. The issue of primary stimulus generalization seems to be negated by experimental evidence bearing on the nature of the generalization mechanism and by appropriate methodological procedures.

The hypothesized effects due to stimulus novelty can be experimentally tested by comparing the partial change design (with novelty effects uncontrolled) and the total change design (with novelty effects controlled) with regard to ID and ED shifts. The following experiment tests for the hypothesized novelty effects.

Dimensional saliency was also included as a factor since the results of House and Zeaman (1962) and Heal, George, and Bransky (1970) suggested that preferred or salient dimensions may be affected by novelty, whereas other dimensions may not. Though House and Zeaman made their suggestion on the basis of experimental data, it is contrary to predictions deduced from their attentional theory of discrimination learning (Zeaman & House, 1963). It would seem that the probability of observing a salient dimension is always close to 1.00, and therefore, the observing probability is so high that novelty would have little or no effect in the postshift phase. That is, there would be a ceiling effect. For low salient dimensions, i.e., for dimensions with small observing probabilities, this would not be the case and novelty might affect the observing probabilities of the low salient dimensions. The probability of observing a dimension could be raised by preexperimental experience or by preshift learning during the experiment. For ID shifts, since the same dimension is relevant in both the pre- and postshift phases, the probability of observing the relevant dimension during the postshift phase is always high. Therefore, novelty should not have an effect on ID shifts. In contrast, preshift training does not increase the observing probability of the relevant dimension used during the ED postshift task. As a result, novelty can facilitate an ED shift when the shift involves a dimension with low saliency.

A note should be made of the fact that, compared to other studies using the partial change design, this study introduces a procedural change. In the other studies, ID shifts were from cues X_1 and X_2 on some relevant dimension X to novel cues X_1' and X_2' on relevant dimension X. Extradimensional shifts were from cues X_1 and X_2 on relevant dimension X to some nonnovel cues Y_1 and Y_2 on relevant dimension Y. That is, in the postshift phase of the ID shift condition, the shift was to the novel cues so that any novelty effects would facilitate ID shifts. But in the postshift phase of the ED condition, the shift was not to the novel cues, so that any novelty effects would hinder ED shifts. In this experiment, ID shifts are the same as above, but ED shifts are from cues Y_1 and Y_2 on relevant dimension Y to novel cues X_1' and X_2' on relevant dimension X. In other words, for both ID and ED shifts, the shift is to the novel cues. Any novelty effects could facilitate both types of shifts. This modification in procedure was introduced so that all *S*s in all conditions sorted the same stimuli using the same relevant dimension and cues during the postshift task.

In summary, a consideration of Slamecka's (1968) methodological arguments leads to the prediction that both ID and ED shifts will be affected by novelty. Slamecka considered novelty to be present whenever the partial change design was used (he did not consider dimensional saliency), such that novelty facilitated ID shifts and hindered ED shifts in the experiments he reviewed. In the present experiment, because of procedural modifications, this line of reasoning leads to the prediction that both ID and ED shifts will be facilitated by novelty. Also, a consideration of the data of House and Zeaman (1962) and Heal et al. (1970) leads to the prediction that novelty will have an effect only when dimensional salience is high. Finally, a consideration of Zeaman and House's (1963) attentional theory leads to the prediction that novelty will have an effect only when dimensional salience is low. This would be the case for an ED shift in which a low salient dimension becomes relevant.

METHOD

Subjects. One hundred twenty students, 54 males and 66 females, from introductory psychology courses at the State University of New York College at Cortland served as *S*s. Students received one point toward their grades for participating.

Material. Four decks of cards, each consisting of 16 3 × 5 cards with different geometrical designs on them were used. For convenience, the decks were labeled Deck 1, Deck 2, Deck 3, and Deck 4. All the decks varied along the five dimensions of form, number of forms on a card, number of dots within each form on a card, size of the forms, and position of the forms on a card. Specifically, the cues of the five dimensions for Deck 4 were circles or squares, one or two forms, three or four dots within each form on a card, 1¼- or ¾-in. forms, and alignment of the forms either in the center of the card in a horizontal row or along either the right or left edge of the card in a vertical column. As an example, one card from Deck 4 consisted of two 1¼-in. squares positioned horizontally in the center of the card with each square containing three dots. Deck 1 was exactly like Deck 4 except that the cues of the position dimension were alignment of the forms either in a vertical column in the center of the card or in a horizontal row along either the top or bottom edge of the card. Deck 3 and Deck 4 had different cues for all dimensions. For Deck 3, the cues were hexagons or triangles, three or four forms, one or two dots within each form on the card, ½- or ¼-in. forms, and alignment of the forms either in the center of the card in a vertical column or in a horizontal row along either the top or bottom edge of the card. All decks were constructed so that the dimensions were orthogonal.

Procedure. Upon entering the laboratory, each *S* was seated at a table and instructed to sort the deck of cards in front of him into two categories labeled Category A and Category B. To help control for any possible effects of primary stimulus generalization, both values of the relevant dimension were alternatively paired with both category labels in the preshift learning and in the postshift phase. This yielded four different combinations of values and labels. Since there were 15 *S*s per cell, all but one of the combinations were used for four *S*s per cell; the other was used for three *S*s. The combination that was used for three *S*s was randomly chosen for each cell. The *S*s were further instructed to turn one card over at a time and to place it in the

TABLE 1

Mean Number of Transformed Trials to Criterion for All Conditions in the Preshift and Postshift Tasks

Tasks	Independent variables							
	Novelty (partial change)				No novelty (total change)			
	High saliency		Low saliency		High saliency		Low saliency	
	ID	ED	ID	ED	ID	ED	ID	ED
Preshift	4.61	3.64	6.74	6.20	4.86	4.25	6.40	5.75
Postshift	3.80	3.95	4.57	5.81	3.63	4.15	4.03	7.84

Note. ID = intradimensional shift; ED = extradimensional shift.

category where they thought it might belong. The *E* told *S* whether he was right or wrong. The *S*s were not allowed to look back through the cards which they had previously sorted. After the instructions were read, *S*s then sorted either Deck 1, 2, or 3 depending upon their cell assignments. After reaching a criterion of 10 consecutive correct responses in the preshift phase, all *S*s were then given Deck 4 and asked to sort it. Criterion learning in the postshift phase was also 10 consecutive responses. If an *S* failed to reach criterion within 160 trials in the preshift phase, he was eliminated from the experiment. Four *S*s were so eliminated; all were in the low salient dimension condition. Since *S*s were eliminated from the low salient dimension condition only, a bias may have been introduced by this differential selection of *S*s. The *S*s of low ability were eliminated from the low salient dimension condition, perhaps raising the average ability of the group in question. However, it should be noted that this selection bias worked against the experimental hypothesis since low salient relevant dimensions should result in a greater number of trials to criterion, whereas higher ability of *S*s should result in fewer trials to criterion.

Since during the preshift learning phase *S*s sorted three different decks, it was necessary to collect data concerning the comparability of the three decks. A total of 57 *S*s were run in order to compare Decks 1 and 3, when either form or number of dots within each form was the relevant dimension, and to compare Decks 2 and 3, when either position of the forms on the card or size of the forms was the relevant dimension. No differences were found. A further test of the equivalence of the decks was provided by the results from the preshift learning phase. Differences would have been indicated by interactions among the conditions.

A second preliminary study was run to scale the saliency of the dimensions used in Deck 4. A total of 36 *S*s sorted Deck 4 with one of the five dimensions relevant. Form was found to be the most salient dimension (median number of trials to criterion = 12), followed by number of forms (median = 14), number of dots within each form (median = 16), size of the forms (median = 40), and position of the forms on the cards (median = 135).

The design of the study was a 2 × 2 × 2 (Novelty × Saliency × Type of Shift) completely randomized design, with 15 *S*s randomly assigned to each of the eight cells. The partial change design was used in the novelty condition and the total change design was used in the no-novelty condition. In the novelty (partial change design) condition, *S*s first sorted either Deck 1 or Deck 2 depending upon whether they were in the high or low salient dimension condition, respectively. In the postshift task, all *S*s in the novelty (partial change design) condition sorted Deck 4 with either forms as the relevant dimension (when *S*s were in the high salient dimension condition) or position as the relevant dimension (when *S*s were in the low salient dimension condition). In the ED conditions, *S*s first sorted either with number of dots within forms as the relevant dimension (when *S*s were in the high salient dimension condition) or with size as the relevant dimension (when *S*s were in the low salient dimension condition).

In the no-novelty (total change design) condition, all *S*s first sorted Deck 3 and then sorted Deck 4. Otherwise, the relevant dimensions for the cells in this half of the experiment corresponded to those used in the analogous cells in the novelty half of the experimental design.

Results

The dependent variables were number of trials to criterion and number of errors to criterion. Analyses of transformed $(x + 1)^{\frac{1}{2}}$ scores for both variables yielded exactly the same results. Therefore, only the transformed number of trials to criterion data are reported. The mean number of transformed trials to criterion for each cell appears in Table 1. In the preshift learning phase, the only significant *F* ratio was due to the main effect of saliency, $F(1, 112)$

= 28.18. (All reported *F* ratios are significant at the .05 level.) In the postshift phase, there were main effects due to saliency, *F*(1, 112) = 22.07, and type of shift, *F*(1, 112) = 15.99. There were also significant interaction effects due to Novelty × Type of Shift, *F*(1, 112) = 4.23, and Saliency × Type of Shift, *F*(1, 112) = 9.32. A posteriori Newman-Keuls tests (using the .05 level of significance) indicated that the no-novelty-ED shift group took a greater number of trials to criterion than the novelty-ID, novelty-ED, and no-novelty-ID shift groups, which did not differ from each other. The respective means of the transformed number of trials to criterion were 6.00, 4.19, 4.88, and 3.83. Similarly, the low-saliency-ED shift group took a greater number of trials to criterion than the high-saliency-ID, high-saliency-ED, and low-saliency-ID shift groups. The respective means were 6.82, 3.71, 4.05, and 4.30. The last three groups were not significantly different from each other.

DISCUSSION

The results of this experiment indicate that there are novelty effects when the partial change design is used and that the novelty effects occur only with ED shifts. This interpretation is based upon the significant Novelty × Type of Shift interaction in which novelty facilitated an ED shift when the shift was to a dimension with novel cues. There was no difference between the novelty-ID and no-novelty-ID shift conditions. Similar effects were observed regarding saliency and type of shift. High salient dimensions facilitated ED shifts, but not ID shifts. The fact that both novelty and dimensional saliency affected the dependent variable in the same way may be an indication that cue novelty increases dimensional saliency.

The findings regarding novelty and dimensional saliency are consistent with Zeaman and House's (1963) attentional theory, and partially justify Slamecka's (1968) criticism regarding the effects of novelty in the partial change design. The results of this experiment are inconsistent with the suggestion based upon the data of House and Zeaman (1962) and Heal et al. (1970) that novelty would only affect highly salient dimensions. The inconsistencies between this study and the two studies mentioned above might be attributable to several differences in the experiments. House and Zeaman and Heal et al. used retarded rather than normal *S*s, and they did not systematically vary novelty, but used the concept in a post hoc explanation of unpredicted results.

Results which might warrant some explanation are the facts that there were no statistical differences between ID and ED shifts under the novelty (partial change design) and high salient dimension conditions (though one should note that the direction of the differences in all cases indicates fewer trials to criterion for the ID shift). The lack of a statistical difference can be explained if one recalls that in the novelty (partial change design) condition, the ED shift was to the dimension with the novel cues, unlike in other studies summarized in Slamecka's (1968) review. In other words, the procedural modification which was introduced into this study to eliminate extraneous variables was responsible for the type of shift result under the condition of novelty. Because of that modification, if novelty had any effect at all, it could only facilitate ED shifts, thereby reducing the magnitude of the usually found difference between ID and ED shifts and yielding a nonsignificant statistical difference. In studies which did not employ the above mentioned procedural modification, novelty would have hindered an ED shift since the shift was away from the dimension with novel cues, thus increasing the difference between ED and ID shifts.

High saliency may have had an effect similar to novelty. In the ID shift condition, the preshift learning raised the probability of observing the relevant dimension close to 1.00, regardless of the degree of saliency. The probability of observing the dimension which was relevant in the preshift learning remained close to 1.00 at the beginning of the postshift phase, and consequently, nothing could possibly raise the probability any higher. Therefore, regardless of whether there were or were not novel cues for this dimension, *S* tended to choose it first, and in the ID condition, received positive feedback, thereby eliminating further search for the relevant dimension. In the ED shift condition, at the conclusion of preshift learning, the relevant dimension was also observed with a probability approximating 1.00, and this probability also remained close to 1.00 at the beginning of the postshift phase. Therefore, *S* continued to respond to the once relevant dimension until he got negative feedback regarding that dimension. After receiving negative feedback, *S*'s probability of observing the once relevant dimension approached zero (see Levine, 1966), and the saliency of the other dimensions then determined which one *S* looked at next. If the relevant dimension was highly salient, *S* looked at it sooner and took fewer trials to criterion. If the relevant dimension was of low salience, *S* looked at it after a greater number of trials.

The conclusion, then, based upon the results of this study, is that *Es* interested in shift learning should control for extraneous novelty effects by using a total change design. Furthermore, previous studies which have used a partial change design should be read critically since novelty effects were confounded with shift effects.

REFERENCES

Graf, V., & Tighe, T. Subproblem analysis of discriminaton shift learning in the turtle (*Chrysemys picta picta*). *Psychonomic Science*, 1971, **25**, 257–259.

Heal, L. W., George, R. T., & Bransky, M. L. The role of cue novelty and dimension dominance in the discrimination shifts of retardates. *Psychonomic Science*, 1970, **18**, 107–108.

House, B., & Zeaman, D. Reversal and nonreversal shifts in discrimination learning in retardates. *Journal of Experimental Psychology*, 1962, **63**, 444–451.

Isaacs, I. D., & Duncan, C. P. Reversal and nonreversal shifts within and between dimensions in concept formation. *Journal of Experimental Psychology*, 1962, **64**, 580–585.

Jeffrey, W. E. Variables affecting reversal-shifts in young children. *American Journal of Psychology*, 1965, **78**, 589–595.

Kalish, H. I. Stimulus generalization. In M. H. Marx (Ed.), *Learning: Processes*. New York: Macmillan, 1969.

Levine, M. Hypothesis behavior by humans during discrimination learning. *Journal of Experimental Psychology*, 1966, **71**, 331–338.

Slamecka, N. J. A methodological analysis of shift paradigms in human discrimination learning. *Psychological Bulletin*, 1968, **69**, 423–438.

Slamecka, N. J. An improved discrimination-shift design. *Psychonomic Science*, 1969, **17**, 89.

Tighe, T., & Frey, K. Subproblem analysis of discrimination shift learning in the rat. *Psychonomic Science*, 1972, **28**, 129–133.

Tighe, T., Glick, J., & Cole, M. Subproblem analysis of discrimination-shift learning. *Psychonomic Science*, 1971, **24**, 159–160.

Zeaman, D., & House, B. The role of attention in retardate discrimination learning. In N. R. Ellis (Ed.), *Handbook of mental deficiency*. New York: McGraw-Hill, 1963.

(Received June 29, 1972)

Journal of Experimental Psychology
1973, Vol. 98, No. 2, 271–278

CHOICES AMONG BETS BY LAS VEGAS GAMBLERS: ABSOLUTE AND CONTEXTUAL EFFECTS[1]

DENNIS G. FRYBACK,[2] BARBARA C. GOODMAN, AND WARD EDWARDS

University of Michigan

When the expected value of a set of gambles is constant, *S*s' choices among those gambles have commonly been interpreted to reflect probability and variance preferences. The first study reported here showed that earlier probability preference findings are replicable in a quite different (casino) setting with substantial stakes. However, the effect of variance was more related to the range of variance offered to *S*s than to absolute levels of variance. Experiment II extended this latter result, producing systematic changes in choice tendencies by manipulating the embedding range of variance within which choices were made. The implication is that the desirability of a gamble is not solely dependent upon the characteristics of that gamble alone.

Decisions are often modeled as choices among gambles of the form "win *A* with probability p_w, otherwise win *B*." The amount *A* is typically positive and *B*, negative or zero. Several variables are available for study. These include the amounts *A* and *B*, the probability of winning, p_w, and two combinations of these: the expected value (EV) and the variance. Typically EV is highly correlated with preferences among gambles (Edwards, 1954b, 1961).

When EV is held constant, various effects of the other variables have been reported. Edwards (1953, 1954a) and Coombs and Pruitt (1960) present data consistent with the hypothesis that *S*s have definite preferences for some levels of p_w and variance over others. Most notably, bets with $p_w = \frac{1}{2}$ were very popular. Lichtenstein (1967) found *S*s preferred lower variance bets. O'Connor, Peterson, and Palmer (1972) report that *S*s tend to hold a variancelike measure constant. However, Slovic and Lichtenstein (1968a) argue that variance, per se, has but a slight effect on gambling decisions and "variance preferences" may be a misnomer for behavior better explained by "strategies that employ only the stated probabilities and payoffs [p. 654]" or subjective transformations of them.

Coombs and Huang (1970) propose that risk is an important determinant in gambling decisions. They assume preference to be monotone increasing in EV and single peaked over risk at each level of EV. In addition, they assume that the risk of playing a gamble *N* times is not less than that of playing that gamble once only. Risk is left undefined. However, it is assumed monotone with variance if p_w is held constant. Data supporting this theory have been reported (Coombs & Huang, 1970; Coombs & Meyer, 1969).

Methodological criticisms cast doubt on these studies. Often *S*s' payment is not contingent upon the outcomes of the gambles chosen; or, if wages are so determined, the stakes are small—typically less than $.25. Slovic, Lichtenstein, and Edwards (1965) demonstrated that such payoff arrangements in gambling experiments generally are boring to *S*s, and that bored *S*s tend to adopt simplistic, systematic strategies not representative of their true gambling preferences. The observed effects of p_w, variance, and the number of times a gamble is played (*N*) may depend on the use of trivial or hypothetical stakes in university-setting experiments.

---

[1] This research was supported by the Wood Kalb Foundation and by U.S. Public Health Service Training Grant GM-01231-08 from the National Institutes of Health to the University of Michigan. Space and facilities were made available by the Four Queens Hotel and Casino in Las Vegas, Nevada. The data were collected by John Ponticello, professional dealer and supervisor at the Four Queens Hotel and Casino.

[2] Requests for reprints should be sent to Dennis G. Fryback, Engineering Psychology Laboratory, H.S.R.I., 2901 Baxter Road, Ann Arbor, Michigan 48105.

In order to study gambling decisions where stakes are real and substantial, a research project was initiated in the Four Queens Casino in Las Vegas. Special experimental games (including the two experiments reported here) were implemented in the casino. The *S*s were any customers of the casino who wanted to play; stakes were comparable to those regularly encountered in the casino's standard games. Payment of winnings and losses was conducted in the standard casino manner. These conditions seem to meet most criticisms that have been directed at the artificiality of previous gambling experiments (Edwards, 1961; Slovic et al., 1965).

EXPERIMENT I

The first experiment tested the hypothesis that earlier findings of probability and variance preferences occur only under trivial payoff conditions. Accordingly, when substantial stakes are employed, distribution of choice frequencies over p_w may not exhibit the sharp peaks normally observed at certain levels of p_w. Variance preference findings may also be tested across *S*s. If each *S* has a preferred level of variance, then two independent random samples of n *S*s from the same population should yield increasingly similar distributions of preferred variance levels for the two groups as n increases. Thus, it is possible to test for variance preferences by offering one group of *S*s gambles in one range of variance, then offering another group of *S*s (independently randomly sampled from the same population as the first group) gambles in a substantially different range of variance, overlapping the first range only in those levels of variance most frequently chosen by the first group. If the two distributions are substantially different when compared in the range of variance common to both sets of gambles, then the notion of absolute variance preferences may be rejected.

Method

Design. The game consisted of choosing 1 bet from a list of 11 bets and then playing that bet. This was repeated for each of nine lists. A second pass through the nine lists was made, but this time *S* played each bet three times (this manipulation of *N* provided a test of the theory of Coombs and Huang, 1970). The game terminated after the second pass.

The *S*s were randomly divided into two groups. Group 1 *S*s made choices from nine Range 1 lists. Group 2 *S*s made choices from nine Range $\frac{1}{3}$ lists identical to Range 1 lists except that all money amounts were divided by three. Thus, the variance of each bet in Range $\frac{1}{3}$ was $\frac{1}{9}$ that of the corresponding bet in Range 1.

The lists of Range 1 are displayed in Table 1. These lists are formed by keeping, in turn, *A*, *B*, and p_w constant; *A* assumes constant values of \$3.00, \$6.00, or \$9.00; *B* assumes constant values of −\$3.00, −\$6.00, or −\$9.00; and p_w assumes constant values of 2/12, 6/12, or 10/12. When *A* (*B*) is constant, p_w varies from 1/12 to 11/12 in steps of 1/12, and *B* (*A*) is varied so that each gamble has EV zero. When p_w is constant, *A* and *B* assume values consistent in range with the other lists. The lists are numbered from 1 to 9, indicating the order in which they were displayed to each *S*.

Subjects. The *S*s were any patrons of the casino who volunteered to play the game. They understood that they would be required to cover all losses using chips purchased from the casino and that they would be paid all winnings in the same manner. The *S*s could quit at any time, but were encouraged to play the entire game. There were 41 *S*s in Group 1 and 34 *S*s in Group 2. Most *S*s won or lost between −\$30.00 and +\$30.00; the extremes were approximately −\$50.00 and +\$200.00.

Procedure. Each *S* was run individually. The *S* was allowed to play as many practice bets as he wished in order to familiarize himself with the game. The game then began for real, and from that point on, all wins and losses were paid according to standard casino practices.

A list of bets was displayed on a cathode ray tube by a PDP-7 computer used for displaying bets and storing data. Each list was displayed with bets arranged so that variance increased from the top to bottom of the list. The column of amounts to win was labeled "WIN," the column of amounts to lose was labeled "LOSE." The p_w was displayed as two numbers for each bet: the number of roulette numbers *S* could pick to win, and the number that were left on which *S* would lose. All bets were played on a standard roulette wheel; if 0 or 00 was spun, the wheel was respun. Thus, 36 numbers were in play. If, for example, p_w was $\frac{1}{12}$, then *S* picked 3 favorable and 33 unfavorable numbers. Once *S* chose a bet he placed markers on the numbers he picked to win; if one of those numbers was spun, *S* won *A*, otherwise he lost *B*.

Once the bet was played, winnings or losses were paid. The computer then displayed the next list until all nine were completed. This procedure was the same for both passes except that each bet was played three times on the second pass.

TABLE 1

THE NINE BASIC LISTS OF BETS

Variable level	Independent variables								
	A			*B*			p_w		
	A	p_w	$-B$	*A*	p_w	$-B$	*A*	p_w	$-B$
	List 5			List 9			List 2		
Low	$3.00	1/12	$ .27	$33.00	1/12	$3.00	$ 3.00	2/12	$ .60
	3.00	2/12	.60	15.00	2/12	3.00	6.00	2/12	1.20
	3.00	3/12	1.00	9.00	3/12	3.00	9.00	2/12	1.80
	3.00	4/12	1.50	6.00	4/12	3.00	12.00	2/12	2.40
	3.00	5/12	2.14	4.20	5/12	3.00	15.00	2/12	3.00
	3.00	6/12	3.00	3.00	6/12	3.00	18.00	2/12	3.60
	3.00	7/12	4.20	2.14	7/12	3.00	21.00	2/12	4.20
	3.00	8/12	6.00	1.50	8/12	3.00	24.00	2/12	4.80
	3.00	9/12	9.00	1.00	9/12	3.00	27.00	2/12	5.40
	3.00	10/12	15.00	.60	10/12	3.00	30.00	2/12	6.00
	3.00	11/12	33.00	.27	11/12	3.00	33.00	2/12	6.60
	List 7			List 1			List 4		
Medium	6.00	1/12	.55	66.00	1/12	6.00	.60	6/12	.60
	6.00	2/12	1.20	30.00	2/12	6.00	3.00	6/12	3.00
	6.00	3/12	2.00	18.00	3/12	6.00	6.00	6/12	6.00
	6.00	4/12	3.00	12.00	4/12	6.00	9.00	6/12	9.00
	6.00	5/12	4.29	8.40	5/12	6.00	12.00	6/12	12.00
	6.00	6/12	6.00	6.00	6/12	6.00	15.00	6/12	15.00
	6.00	7/12	8.40	4.29	7/12	6.00	18.00	6/12	18.00
	6.00	8/12	12.00	3.00	8/12	6.00	21.00	6/12	21.00
	6.00	9/12	18.00	2.00	9/12	6.00	24.00	6/12	24.00
	6.00	10/12	30.00	1.20	10/12	6.00	27.00	6/12	27.00
	6.00	11/12	66.00	.55	11/12	6.00	30.00	6/12	30.00
	List 3			List 6			List 8		
High	9.00	1/12	.82	99.00	1/12	9.00	.60	10/12	3.00
	9.00	2/12	1.80	45.00	2/12	9.00	1.20	10/12	6.00
	9.00	3/12	3.00	27.00	3/12	9.00	1.80	10/12	9.00
	9.00	4/12	4.50	18.00	4/12	9.00	2.40	10/12	12.00
	9.00	5/12	6.43	12.60	5/12	9.00	3.00	10/12	15.00
	9.00	6/12	9.00	9.00	6/12	9.00	3.60	10/12	18.00
	9.00	7/12	12.60	6.43	7/12	9.00	4.20	10/12	21.00
	9.00	8/12	18.00	4.50	8/12	9.00	4.80	10/12	24.00
	9.00	9/12	27.00	3.00	9/12	9.00	5.40	10/12	27.00
	9.00	10/12	45.00	1.80	10/12	9.00	6.00	10/12	30.00
	9.00	11/12	99.00	.82	11/12	9.00	6.60	10/12	33.00

Results

Figures 1a and 1b show relative frequency distributions over the levels of p_w for lists in which *A* and *B* were constant, respectively, for $N = 1$. The solid and dotted lines display results for Groups 1 and 2, respectively. Figures 1c and 1d are comparable plots for $N = 3$. All four plots show a marked peak in relative frequency of choices at $p_w = \frac{1}{2}$; these peaks are most prominent when $N = 3$. The four probability levels $\frac{1}{4}$, $\frac{1}{3}$, $\frac{1}{2}$ and $\frac{2}{3}$ account for almost 60% of all choices by *Ss* in either group while these levels comprise but 33% of possible choices (excluding lists with p_w constant). The effect of varying *N* seems to be that of slightly strengthening the tendency to choose a bet with $p_w = \frac{1}{2}$ when $N = 3$ as opposed to $N = 1$.

The effect of *N* on variance is just as

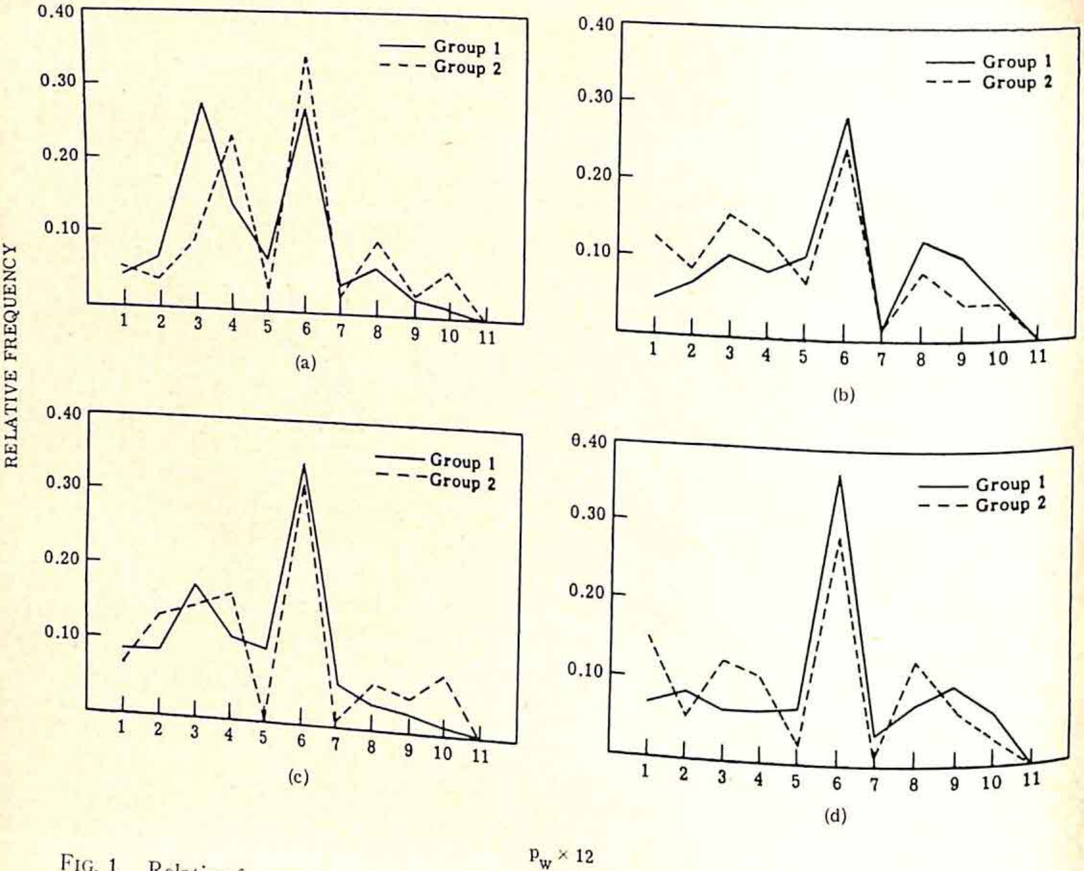


FIG. 1. Relative frequencies with which the different p_{w}s were played. (For Groups 1 and 2, respectively, the distributions are based on (*a*) 147, 132, (*b*) 145, 129, (*c*) 129, 118, (*d*) 130, and 112 observations.)

slight. For each list the variance of the bet chosen by each *S* when $N = 1$ was compared to the variance of the bet that *S* chose when $N = 3$. In general, no consistent within-*S* trends were found. The variance of the bet chosen when $N = 3$ could have been larger than, equal to, or smaller than the variance of the bet chosen when $N = 1$. Most *S*s tended to have their comparison choices almost equally divided among these three possibilities irrespective of whether A, B, or p_w was the constant parameter of the list. However, when an *S* increased or decreased variance it was generally the minimal change possible from the variance of the $N = 1$ choice.

By far the most interesting result was obtained in the examination of the between-group test for variance preferences. The idea that *S*s maintain preferences for absolute levels of variance can be rejected on the basis of these data. The *S*s in Group 2 tended to play at different absolute levels of variance than did Group 1 *S*s in spite of the fact that at any time they could have picked comparable levels and vice versa. Yet the choice behavior of the two groups is very similar with respect to relative variance. Figure 2, a plot of cumulative proportions of choices against within-range rank order of variance levels, displays this similarity. (Note that only data for $N = 1$ are included in Figure 2; the $N = 3$ data are nearly identical and are not displayed.) If Group 2 (1) *S*s had been attempting to play absolute variance levels comparable to Group 1 (2), then the dashed (solid) curve should have appeared in the right- (left-) hand third of the graph.

Figure 2 shows data pooled across all nine lists for each group; however the same degree of similarity is seen in comparable analyses for data from the constant p_w lists only. Although the overall similarity is striking, the difference between groups is significant statistically ($p < .05$, two-tailed Kolmogorov-Smirnov test).

Discussion

Perhaps the least interesting result of Experiment I is the fact that earlier findings of probability preferences can be replicated utilizing gambles with substantial monetary outcomes. Manipulating N seemed to have little effect. The theory of Coombs and Huang (1970) predicts that, when p_w is constant, each S should pick a bet to play three consecutive times that is equal to or lower in variance than the bet he picked from the same list to play once only. This prediction was not supported in this real gambling environment; each S tended to increase variance on about a third of the lists irrespective of the constant parameter in the list. However, such increases in variance were generally the minimum possible.

The interesting finding is that regarding the effect of variance. It is clear that *S*s do not seem to have preferences for absolute levels of variance. Rather, the effect of variance seems to be mediated by the range of variances within which *S*s make their choices, contradicting previous notions of variance preferences. It is true that the difference in relative choice behavior of the two groups was significant statistically, but this difference seems unimportant in the face of the low a priori expectation of the high degree of similarity between the two groups. The small difference between the relative behavior of the two groups probably represents a slight effect of variance in absolute terms; *S*s in the low-variance range tended to pick bets slightly higher in variance (relatively) than did *S*s in the higher range. Yet there is a high degree of similarity between the two groups. A reasonable explanation for this similarity is that variance preferences exist, but are more of a relative than absolute nature. The *S*s seem to be more influenced by whether a gamble is low, medium, or high with respect to the set of gambles determining the context in which choices are being made (producing the similarity between the groups) than by the absolute level of variance of the gamble (representing the difference between the two groups).

Four other possible explanations for the similarity between the two groups were examined. Strategies based on choosing bets equal in p_w, A, or B could not account for the data. The high degree of similarity between groups was observed within the lists for which each of these three variables, in turn, was held constant. Strategies such as "always pick Bet 3" were not observed and could not account for the degree of similarity observed, even if they had occurred.

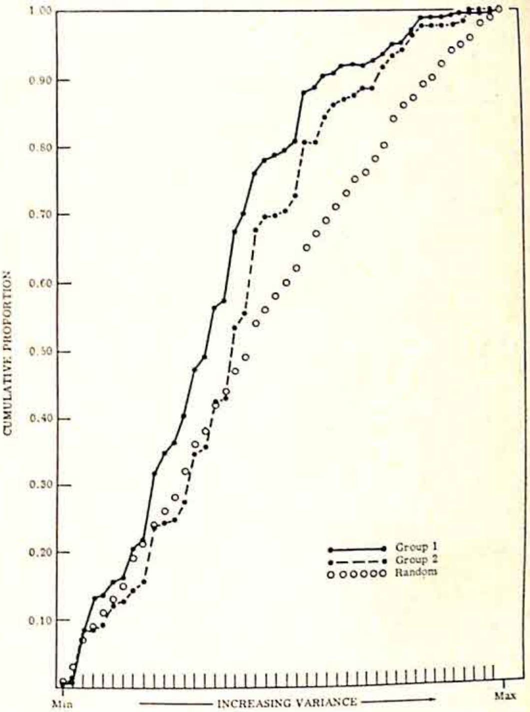


FIG. 2. Cumulative proportions of choices for Groups 1 and 2 and theoretical random *S*s over within-range ordinal levels of variance.

Experiment II was designed to test the relative variance explanation for similarity in relative choice behavior observed in Experiment I.

EXPERIMENT II

The gambles in Range 1 of Experiment I were on the whole relatively high-variance gambles when compared to Range $\frac{1}{3}$ gambles and vice versa. We hypothesized that if a group of *S*s who had first played bets from Range 1 were then switched to Range $\frac{1}{3}$ lists, they would pick relatively higher variance bets in Range $\frac{1}{3}$ than they would have done if they had been playing in Range $\frac{1}{3}$ all along. Similarly, *S*s playing

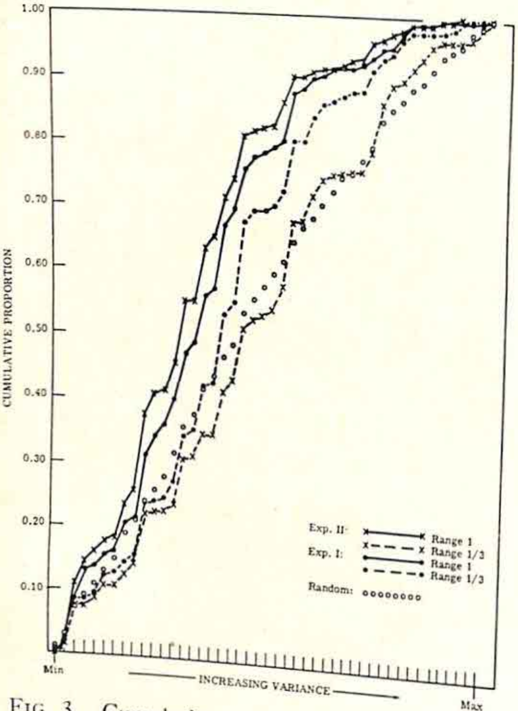


FIG. 3. Cumulative proportions of choices over within-range ordinal levels of variance for Group 1 compared to Experiment I results and random strategy.

in Range $\frac{1}{3}$ when switched to Range 1 should play lower variance bets than they would have otherwise. What if lists from the two ranges were intermixed? Since most variance levels in Range $\frac{1}{3}$ were also available in the higher range, we expected that the overall context would be that of the high range. Thus, the effect should be the same as that produced by having *S*s first play in Range 1 and then switch to Range $\frac{1}{3}$—they would pick relatively higher variance bets in Range $\frac{1}{3}$ than if presented those lists alone.

These three manipulations form the basis of Experiment II.

Method

Design. The *S*s were divided into three groups corresponding to three presentation orderings of the lists of bets used in Experiment I. Group 1 played a block of 12 lists in Range 1, then a block of 12 lists in Range $\frac{1}{3}$; Group 2 played the 24 lists in a fixed randomized order; Group 3 played 12 lists from Range $\frac{1}{3}$, then 12 lists from Range 1.

The first nine lists in each block were the nine basic lists used in Experiment I. The last three lists in the block were formed from Lists 4, 5, and 9 by deleting the bet chosen by *S* when those lists were previously presented. Thus, if *S* chose Bet 5 from List 4, then the tenth list he saw was comprised of Bets 1–4 and 6–11 from List 4. This was done to examine *S*s' "second-choice" bets from the lists. All *S*s played three more lists. These last three lists were Lists 4, 7, and 1 from the first block of 12 lists that *S* saw; however, instead of being displayed in the orderly format of the previously displayed lists, the individual bets were displayed in a scrambled order with respect to variance. Thus, every *S* completing the game played 27 bets.

Subjects. The *S*s were again casino patrons who volunteered to play the game with the understanding that the wins and losses incurred would be dealt with exactly as in any other casino game. There were 36, 34, and 6 *S*s in Groups 1, 2, and 3, respectively. The low number of *S*s in Group 3 was due to the fact that the study was forced to terminate prematurely in Las Vegas.

Procedure. The display and playing of the bets were the same as in Experiment I. Data for each condition were collected before the next was implemented.

Results

Although *S*s were encouraged to play all the bets, they could of course quit whenever they desired; three and seven *S*s in Groups 1 and 2, respectively, did so. In addition, three *S*s in Group 2 had previously participated in Group 1. Data for these 13 *S*s were excluded from the analyses. Thus, 33, 24, and 6 data records were used for Groups 1, 2, and 3, respectively.

Figures 3 and 4 show cumulative proportions of choices as a function of relative level of variance for Groups 1 and 3, respectively. Results of Experiment I are included for comparison in both figures. Distributions for Group 2 were essentially identical to those for Group 1 and are not displayed. All statements about Group 1 apply equally to Group 2.

As may be seen in Figure 3, Group 1, with prior experience in Range 1, tended to play higher variance levels in Range $\frac{1}{3}$ than would be predicted from Experiment I results. Though the Range 1 distribution for these *S*s differs somewhat from the corresponding distribution from Experiment I, the difference is not significant ($p > .10$, two-tailed Kolmogorov-Smirnov test). However, the difference for Range $\frac{1}{3}$ between Group 1 and the corresponding Experiment I result is highly significant ($p < .001$, one-tailed Kolmogorov-Smirnov test) in the predicted direction. On the basis of the few *S*s that were obtained in Group 3, the data in Figure 4 support the prediction that *S*s playing in the low range then the high range would play lower variance bets in the higher range than otherwise. In this group *S*s tended to replicate the Experiment I result for Range $\frac{1}{3}$ (the difference is not significant at the $p < .10$ level) and were significantly different from Experiment I *S*s when playing the Range 1 lists ($p < .001$, one-tailed Kolmogorov-Smirnov test). Only choices from the nine basic lists in each range were used to compute the distributions in Figures 3 and 4. The second-choice lists and the scrambled lists were analyzed separately.

Results from the second-choice lists are generally inconclusive. Correlations between variance of first-choice bet and variance of the second-choice bet across *S*s varied from −.14 to .75 for the six lists used to collect second choices. In general first-choice variance was a relatively poor predictor of second-choice variance. Lists 1, 4, and 7 were presented twice—once in orderly (with respect to variance) form and once in scrambled form. Correlations between the variances of the bets chosen in the two forms were .33, .37, and .19 for Group 1, and .73, .30, and .31 for Group 2 for the three lists, respectively. There were too few *S*s in Group 3 to perform these analyses. Eighteen *S*s in Groups 1 and 2 played the game more than once. For each list the variance of the bet chosen the first time an *S* played was compared to the variance of the bet he chose the second time he played the game. Across-*S* correlations of these variances for the various lists fell between .16 and .89, with mean correlation across lists of .54. It seems, therefore, that the overall group behavior was quite stable with respect to variance (as evidenced by replication of Experiment I results), but within *S* there was a good deal of variability.

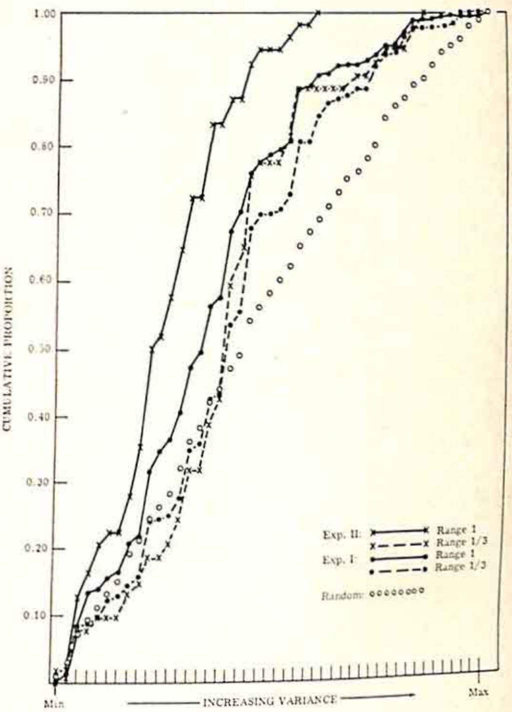


FIG. 4. Cumulative proportions of choices over within-range ordinal levels of variance for Group 3 compared to Experiment I results and random strategy.

GENERAL DISCUSSION

The results of these two experiments are clearly at odds with most theories that attempt to explain choices among gambles. When EV is constant and substantial amounts of money are at stake, the desirability of a gamble is not solely dependent upon the characteristics of that gamble. Experiment I replicated earlier findings of probability preferences; but the results of Experiments I and II together imply that none of the parameters of a gamble, alone or in combination, is sufficient to index the desirability of that gamble in the absolute sense. Evidently the choice probability orderings of a given set of gambles can be changed by embedding that set of gambles in differing contexts. This conclusion stands regardless of whether one accepts the explanation offered in terms of variance.

Specifically, regression models (Slovic & Lichtenstein, 1968b) and subjectively expected utility theory (Edwards, 1954b, 1961), without modification, are clearly inadequate to explain the shifts in choice behavior observed in Experiment II. Portfolio theory (Coombs & Huang, 1970) could account for these results as shifts due to changes in the perceived risks of the gambles; however, this would require modification of the ad hoc assumptions about perceived risk that enable tests of the theory.

More generally, doubt is cast on probabilistic choice theories that assume independence among alternatives when those alternatives are zero-EV gambles. One common form of this assumption is that which states that the choice probability orderings of any two stimuli are independent of the offered set of stimuli (a specific example is the theory of Luce, 1959; Tversky, 1972). Although the present study did not examine choice probabilities directly, the fact that it is possible to change *Ss'* choice tendencies by manipulating the set of offered alternatives strongly suggests that order independence as stated does not obtain.

Recently Tversky (1972) has proposed a probabilistic choice model based on elimination by aspects. This model explicitly does not assume order independence as stated above. However, Tversky derives as a consequence of his model a regularity condition, which asserts that the probability of choosing a given alternative cannot be increased by enlarging the offered set containing that alternative. Although this specific manipulation was not included in the present experiment, on the basis of the results we would expect that it is possible to increase choice tendencies for, say, the higher variance bets in one of the Range $\frac{1}{3}$ lists merely by adding even higher variance bets to the list. This would obviously contradict the regularity condition necessary to Tversky's model.

That the perception of a given stimulus can be influenced by other stimuli and prior experience is not a foreign idea in psychology. Judgmental anchoring and adaptation effects have long been in the literature (e.g., Helson, 1947). Contrast and assimilation are common terms in social psychology (e.g., see Sherif & Hovland, 1961). The present results suggest that concepts analogous to anchoring, adaptation, assimilation, and contrast can no longer be ignored by any theory that attempts to describe human decision making.

REFERENCES

Coombs, C. H., & Huang, L. C. Tests of a portfolio theory of risk preference. *Journal of Experimental Psychology*, 1970, **85**, 23–29.

Coombs, C. H., & Meyer, D. E. Risk-preference in coin toss games. *Journal of Mathematical Psychology*, 1969, **6**, 514–527.

Coombs, C. H., & Pruitt, D. G. Components of risk in decision making: Probability and variance preferences. *Journal of Experimental Psychology*, 1960, **60**, 265–277.

Edwards, W. Probability-preferences in gambling. *American Journal of Psychology*, 1953, **66**, 349–364.

Edwards, W. The reliability of probability preferences. *American Journal of Psychology*, 1954, **67**, 88–95. (a)

Edwards, W. The theory of decision making. *Psychological Bulletin*, 1954, **51**, 370–417. (b)

Edwards, W. Behavioral decision theory. *Annual Review of Psychology*, 1961, **12**, 473–498.

Helson, H. Adaptation-level as frame of reference for prediction of psychophysical data. *American Journal of Psychology*, 1947, **60**, 1–29.

Lichtenstein, S. C. Bases for preferences among three-outcome gambles. *Journal of Experimental Psychology*, 1967, **69**, 162–169.

Luce, R. D. *Individual choice behavior: A theoretical analysis.* New York: Wiley, 1959.

O'Connor, M. F., Peterson, C. R., & Palmer, T. J. Stakes and probabilities in information purchase. *Organizational Behavior and Human Performance*, 1972, **7**, 43–52.

Sherif, M., & Hovland, C. I. *Social judgment: Assimilation and contrast effects in communication and attitude change.* New Haven: Yale University Press, 1961.

Slovic, P., & Lichtenstein, S. C. Importance of variance preferences in gambling decisions. *Journal of Experimental Psychology*, 1968, **78**, 646–654. (a)

Slovic, P., & Lichtenstein, S. C. Relative importance of probabilities and payoffs in risk taking. *Journal of Experimental Psychology*, 1968, **78**, 1–18. (b)

Slovic, P., Lichtenstein, S. C., & Edwards, W. Boredom-induced changes in preferences among bets. *American Journal of Psychology*, 1965, **78**, 208–217.

Tversky, A. Elimination by aspects: A probabilistic theory of choice. *Psychological Review*, 1972, **79**, 281–299.

(Received June 29, 1972)

Journal of Experimental Psychology
1973, Vol. 98, No. 2, 279–285

LATENCY OPERATING CHARACTERISTIC:

III. TEMPORAL UNCERTAINTY EFFECTS[1]

JOSEPH S. LAPPIN[2] AND KENNETH DISCH
Vanderbilt University

The purpose of this study was to determine the effect of temporal uncertainty on the latency operating characteristic (LOC), defined as the trade-off between speed and accuracy. Foreperiods were either constant (1,000 msec.), one of two equally likely but unnoticeably different durations (975 and 1,025 msec.), or one of two equally likely and noticeably different durations (750 and 1,250 msec.). Mean RT increased with temporal uncertainty, but this was not always so for the LOC. Most efficient performance was obtained when the foreperiod was 975 msec., and poorest performance occurred in the 1,250-msec. condition. Better performance with the early stimulus (at 975 msec.) is contrary to Anticipation and Preparatory Set models for the effects of temporal uncertainty, but is consistent with a Diluted Sample model in which temporal uncertainty affects the temporal integration of sensory data.

Reaction time (RT) increases with temporal uncertainty. Increasing either the variability (e.g., Karlin, 1959) or the length (Posner & Boies, 1971; Woodrow, 1914) of the foreperiod between warning signal and stimulus reliably produces longer RTs.

The oldest and most widely accepted explanation for this phenomenon is that RT is inversely related to *S*'s preparedness for perceiving and responding to the stimulus, and this preparedness presumably decreases with *S*'s expectancy for the stimulus presentation. The explanation will be termed the Preparatory Set model. Following the rationale elaborated by Thomas (1967), the *S*'s expectancy will be represented as a hazard function, i.e., the subjective probability that the stimulus will be presented at time t given that it has not yet been presented. For the present experiment, in which the foreperiod is selected with equal probability from a small set of alternatives, the principal characteristic of this representation is that *S*'s expectancy is an increasing function of time from the onset of the trial. Preparation should, thus, be greater for the longer of several alternative foreperiods and should be inversely related to foreperiod variability.

However, temporal uncertainty not only produces longer RTs but also tends to produce lower error rates in choice tasks (Bertelson, 1967) and fewer anticipatory responses in simple RT tasks (Nickerson, 1965; Snodgrass, 1969). These results clearly indicate that various foreperiod conditions yield varying biases for responding within a given range of RT, independent of any effects the foreperiod might have on the perceptual processing per se. Snodgrass (1969) concluded that anticipatory responses triggered by *S*'s estimation of the foreperiod were sufficient to account for foreperiod effects on simple RT. Snodgrass' anticipatory model postulates two discrete sets of RTs, true RTs and anticipations; but a more general alternative might be simply that *S*'s criterion for responding varies with his expectation of the stimulus occurrence, without restricting the information from sensory processes to be binary, multivalued, or continuous. This general explanation for the effects of temporal uncertainty will be designated as the Anticipation model. The essence of this model is that temporal uncertainty affects decision processes but not the sensory encoding of the stimulus nor the production of a motor response.

---

[1] Preparation of this report was partially supported by Grant MH-21105 from the National Institute of Mental Health. The results were also presented at the Midwestern Conference on Mathematical and Theoretical Psychology in Bloomington, Indiana, April 1972.

[2] Requests for reprints should be sent to Joseph S. Lappin, Department of Psychology, Vanderbilt University, Nashville, Tennessee 37240.

A third explanation for temporal uncertainty effects can be derived from assumptions (*a*) that the detection and recognition of a stimulus depend upon an accumulation of sensory data beginning at some ill-defined point in time, (*b*) that *S* can adjust the point in time at which the sampling of sensory data is initiated, and (*c*) that the tendency to begin integrating the incoming sensory data varies monotonically with *S*'s expectation of the stimulus onset. Conditions of temporal uncertainty presumably produce a greater expectancy prior to the stimulus onset, a greater weight for sensory data occurring prior to the stimulus onset, and a concomitant reduction in the signal–noise ratio when the stimulus is presented. This explanation will be referred to as the Diluted Sample model. The essential idea is that the preparatory process is cumulative. Whereas the Preparatory Set model has focused on the lower expectancy at the point of stimulus onset, the Diluted Sample model points to the higher expectancy prior to the stimulus onset. Both models predict poorer performance as the foreperiod is lengthened in a constant-foreperiod experiment, but make different predictions when the foreperiod is selected from a set of equally likely alternatives: The Preparatory Set model predicts better performance for longer foreperiods; the Diluted Sample model predicts better performance with shorter foreperiods. Although many experimenters have found faster RTs for the longer foreperiods in a set of equally likely alternatives (e.g., Snodgrass, 1969), this may be due simply to shifts in the decision criterion. Green and Luce (1971) found a slight tendency for faster RTs with shorter foreperiods when they used a distribution of foreperiods with a constant hazard rate.

Any attempt to explain the effects of temporal uncertainty upon performance must differentiate the contributions of processes for encoding and decoding stimulus information from the contributions of decision processes that bias the *a priori* tendency for responding at a particular point in time. Our rationale for differentiating these component processes is based upon the trade-off relation between speed and accuracy, i.e., the "latency operating characteristic" (LOC) (Lappin & Disch, 1972a, 1972b). Variation in decision criteria is presumed to shift the speed and accuracy to various points on the same LOC, whereas the amount of perceptual information available at any given time is presumably reflected in the discrimination accuracy of responses made at a given RT. The slope and the intercept of a linear LOC provide two parameters for measuring the speed of perceptual processing.

The Anticipation model predicts that the LOC should remain invariant with temporal uncertainty since temporal uncertainty is assumed to affect the response criterion but not the perceptual processing. If various foreperiod conditions were to produce changes in RT distributions but no variation in the LOC, then evidence would be provided for the convergent validity of the LOC as a measure of perceptual processes unaffected by decision processes. Both the Preparatory Set and Diluted Sample models predict a shift in the LOC with temporal uncertainty. Insofar as the contributions of expectancy might be assumed to occur on a probabilistic basis, rather than as a constant additive component in the speed of perceptual processing, then the predicted effects should appear in the slope rather than intercept of the LOC.

Three conditions of temporal uncertainty were employed in this experiment: The foreperiod in various sessions was either (*a*) constant (1,000 msec.), (*b*) one of two equally likely but unnoticeably different durations (975 or 1,025 msec.), or (*c*) one of two equally likely and noticeably different durations (750 and 1,250 msec.). Temporal uncertainty should obviously be greater in the latter condition than in either of the other two; the effect of the second condition with unnoticeably different durations was, however, less certain. This unnoticeable condition was assumed to give rise to only a single distribution of time estimations, whereas the noticeable condition should produce separate estimations of the short and long foreperiods. The unnoticeable condition thus provides for an examination of fine-grain changes in the temporal course of *S*'s information processing. Expectancy is presumed to be greater with the long than with the short foreperiods in both noticeable and unnoticeable conditions.

METHOD

Design. Two groups of *S*s performed under different foreperiod conditions. For Group N, the first four sessions utilized two equally likely and noticeably different foreperiods of 750 and 1,250 msec., and the last two sessions utilized a constant foreperiod of 1,000 msec. For Group U, the first four sessions utilized two equally likely but unnoticeably different foreperiods of 975 and 1,025 msec, and the last two sessions utilized a constant foreperiod of 1,000 msec. The first two sessions for both groups were considered practice, and the data from these were not analyzed. The last two sessions were identical for each group; some pilot data had led us to suspect that the two groups might perform differently as a result of the preceding performance under different conditions of temporal uncertainty although this phenomenon was not obtained in the present experiment.

Subjects. The *S*s were 12 Vanderbilt students who served as paid volunteers. Six *S*s were graduate students, and the remainder were undergraduates. Five were female and seven were male. All *S*s were naive as to the purpose of this experiment, and all had little or no experience with similar RT experiments. Six *S*s were randomly assigned to each group; three graduate students were in each group. One group had three females; the other had two.

Stimuli and apparatus. Most of the methodological details of this experiment were identical to those of the two previously reported experiments (Lappin & Dish, 1972a, 1972b). The stimuli were luminous circles, 2.1 c/m² in luminance, 1° in diameter, and 40′ off center to the right or left of the fixation point, remaining in view until *S* responded. The background was 5.5° square, 20.6 c/m² in luminance. The two alternative responses consisted of lifting the right or left hand from a telegraph key; the assignment of responses to the two stimulus positions was compatible.

Trials were initiated by *S* depressing both of the telegraph keys, and the stimulus was presented after the appropriate foreperiod. Foreperiod durations and stimulus positions were controlled by prepunched paper tape; both variables were independently randomized with the restriction that each of the two values occurred 45 times in each successive block of 90 trials. The stimuli were presented and controlled with the same apparatus (Iconix) as in the previous two experiments.

Procedure. A speed–accuracy trade-off was obtained by asking *S*s to respond at a speed that would result in an error rate of about 25%. The *S*s were instructed to avoid guessing.

Each *S* performed on 392 trials in each of the six sessions. The first 32 trials were considered practice, and their data were not analyzed. There was no break between practice and experimental trials. There were 180 experimental trials at each of the two alternative foreperiods in each of the two experimental sessions with temporal uncertainty and 360 trials per session with the constant 1,000-msec. foreperiod.

As a check on whether the foreperiods of 975 and 1,025 msec. were indeed unnoticeably different, *S*s in Group U were asked at the end of the experiment whether they had ever suspected that the signals arrived at different times on different trials and whether they had noticed any change in the procedure for the last two sessions. None of the six *S*s gave any indication of having noticed a difference in foreperiod between trials or sessions. Our assumption is simply that the distribution of *S* strategies and time estimations was the same for foreperiods of 975 and 1,025 msec. We do not assume that *S*s were not *able* to discriminate the difference in foreperiods, but only that they did not do so. The *S*'s expectancy at 975 msec. was thus assumed to be less than or equal to that at 1,025 msec. after the beginning of the trial.

Additional details of the procedure were identical to the previous two experiments (Lappin & Disch, 1972a, 1972b).

RESULTS

The average speed and accuracy of each *S* with each foreperiod are given in Table 1. In addition to the mean RT and error rate, Table 1 also provides an estimate of the "stimulus-controlled RT" according to the correction-for-guessing formula that follows from Yellott's (1971) "fast-guess model." These statistics are presented primarily for descriptive purposes; a more detailed analysis of performance is given below in terms of the LOC.

As may be seen, mean RT was an increasing function of temporal uncertainty. In Group N, the fastest RTs occurred with the constant 1,000-msec. foreperiod, and the 1,250-msec. foreperiod produced faster RTs than the 750-msec. foreperiod. The effect of foreperiod on mean RT in Group N is statistically significant by the Friedman two-way analysis of variance by ranks, $\chi_r^2 = 10.3$, $n = 6$, $k = 3$, $p < .01$ (see Siegel, 1956). This same effect of temporal uncertainty on mean RT is also reflected by a comparison between the two groups, N and U, for the third and fourth sessions with variable foreperiods: Group U was faster. The average RTs for Group U (average for the 975-msec. and 1,025-msec.

TABLE 1

MEAN RT (IN MSEC.), PROBABILITY OF AN ERROR, AND ESTIMATED "STIMULUS-CONTROLLED RT" (μ_s) FOR EACH SUBJECT AND FOREPERIOD

Foreperiod	Subjects													
	Noticeable							Unnoticeable						
	1	2	3	4	5	6	Average	7	8	9	10	11	12	Average
750														
M	239	226	232	222	222	222	227							
P_e	.194	.214	.206	.192	.181	.214	.200							
μ_s	255	240	248	229	233	236	240							
975														
M								211	181	183	172	196	176	186
P_e								.172	.197	.206	.269	.300	.197	.224
μ_s								235	204	196	206	244	184	212
1,000														
M	218	210	212	200	188	203	205	202	212	178	173	212	203	196
P_e	.240	.239	.235	.169	.271	.251	.234	.206	.215	.271	.328	.221	.296	.256
μ_s	254	234	231	211	234	225	232	228	241	200	211	259	222	227
1,025														
M								193	221	198	189	203	224	205
P_e								.247	.186	.186	.233	.267	.256	.229
μ_s								242	241	207	207	249	240	231
1,250														
M	226	208	227	220	214	206	217							
P_e	.291	.267	.222	.178	.256	.283	.250							
μ_s	270	237	242	227	229	217	237							

Note. $\mu_s = (P_cM_c - P_eM_e)/(P_c - P_e)$, where P_c and P_e are the probabilities of correct and erroneous responses and M_c and M_e are the mean RTs for correct and erroneous responses (see Yellott, 1971).

foreperiods) are significantly faster than those for Group N (average for the 750-msec. and 1,250-msec. foreperiods) by a Kruskal-Wallis one-way analysis of variance by ranks, $\chi^2(1) = 8.3, p < .01$. Foreperiod duration had no significant effect on the mean RTs for Group U, by the Friedman test, although there was an unexpected tendency for the faster RTs to occur with the hortest foreperiod.

However, speed was often associated with increased errors; and when the mean RTs are "corrected for guessing" by Yellott's (1971) fast-guess model, then the correct values for "stimulus-controlled RTs" reveal much smaller effects due to temporal uncertainty. A more detailed analysis of the effect of the foreperiod on the efficiency of performance was based on the LOC.

Figure 1 presents the LOC for each foreperiod condition. Each data point is based on the combined data from all six *S*s in each group; the RT on the abscissa is the median RT for each successive block of 270 trials (except as limited by ties), ranked from fastest to slowest for all six *S*s combined. Various *S*s therefore make slightly different contributions to each data point, depending on their individual distributions of RTs. The data were pooled in this way because the number of trials for each *S* at each foreperiod in the uncertain conditions (360 trials) did not permit a reliable estimate of the LOC. The linear equations by which these LOC functions are described were calculated by the same least-squares procedure employed in our previous reports. The percentages of variance in d' accounted for by these equations were as follows for each of the six conditions: N 750, 90.2%; N 1,250, 98.4%; N 1,000, 94.1%; U 975, 96.2%; U 1,025, 95.9%; U 1,000, 95.8%.

As may be seen, performance tended to be superior under the 975-msec. foreperiod in Group U and less efficient under the uncertain foreperiod conditions (750 and 1,250 msec.) in Group N.

Statistical analyses of the differences among the various LOCs were based on Fischer's exact probability test for comparing two independent samples of nominal data (see Siegel, 1956). In order to compare any two conditions, the first step was to calculate the average LOC for the two conditions (average intercept and average slope) and then tabulate the number of points for each condition in which the discrimination accuracy (d') was above or below this line for just those points with median RT above the average intercept. A 2 × 2 table was compiled by this procedure for each of the 15 pairs of conditions. The results are given in Table 2. The upper right half of this matrix gives the statistical significance of the comparison between each pair of conditions, and the lower left half gives the 2 × 2 table on which each comparison is based. The six conditions seem to fall into three rough categories: most efficient performance under the 975-msec. foreperiod condition, an intermediate and similar level of efficiency for the 1,025-msec. and the two constant 1,000-msec. foreperiods, and poorest performance under the 750- and 1,250-msec. foreperiods.

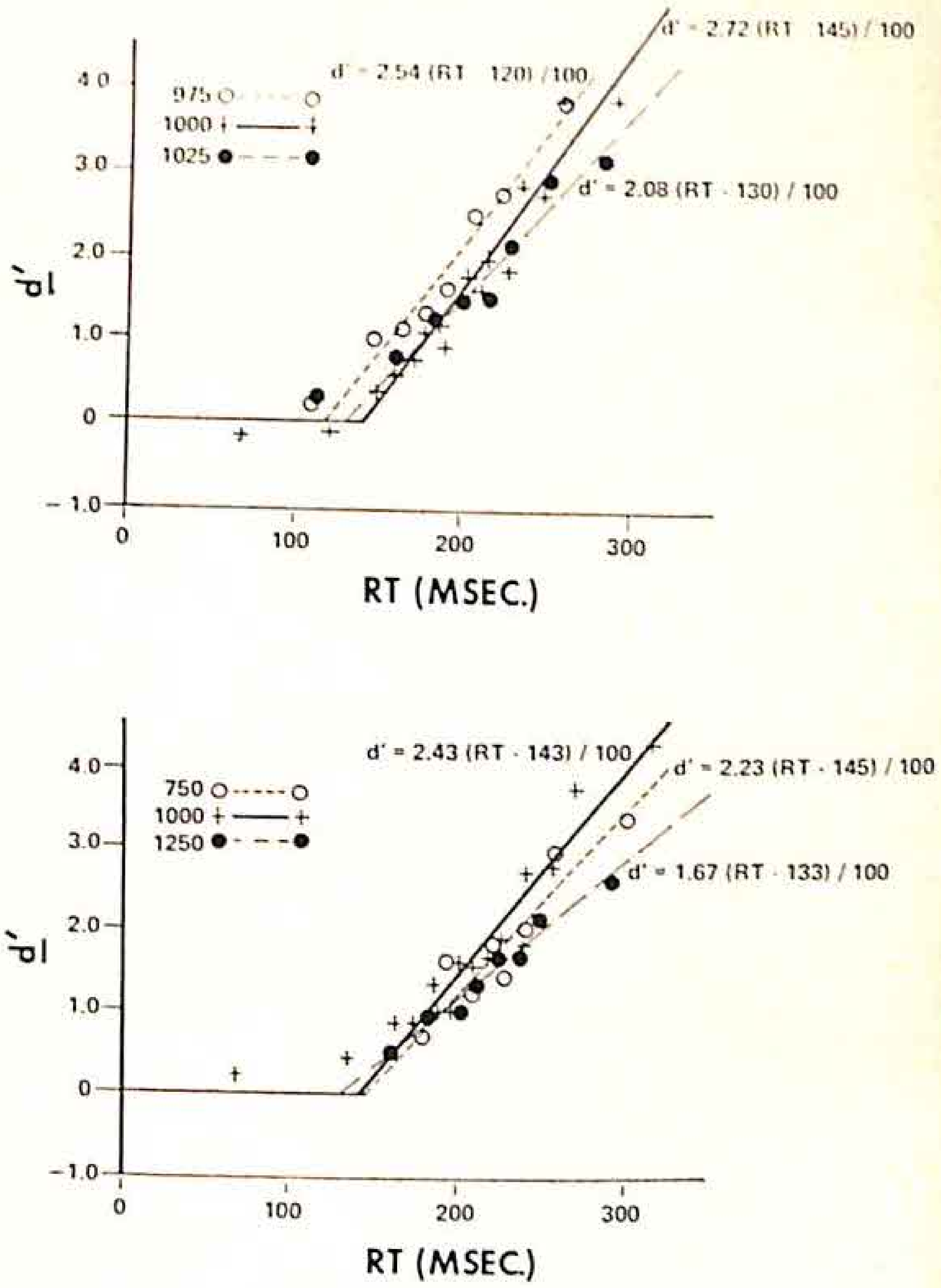


FIG. 1. Latency operating characteristics for the combined data of each group (Group U in the upper section and Group N in the lower section).

TABLE 2

PAIR-WISE COMPARISONS IN DISCRIMINATION ACCURACY (d') FOR EACH FOREPERIOD CONDITION

Condition	1	2	3	4	5	6
N 1,250 (1)		*ns*	$p < .01$	$p < .01$	$p < .01$	$p < .01$
N 750 (2)	5 | 3 2 | 6		*ns*	*ns*	$p < .05$	$p < .01$
N 1,000 (3)	10 | 4 0 | 8	8 | 6 2 | 6		*ns*	*ns*	$p < .01$
U 1,025 (4)	6 | 1 0 | 8	5 | 2 2 | 6	4 | 3 7 | 7		*ns*	$p < .05$
U 1,000 (5)	12 | 2 0 | 8	11 | 3 2 | 6	8 | 6 6 | 8	7 | 7 3 | 4		$p < .01$
U 975 (6)	7 | 0 0 | 8	7 | 0 0 | 8	7 | 0 2 | 12	5 | 2 0 | 7	6 | 1 2 | 12	

Note. Conditions are ranked poorest to best from left to right. In the 2 × 2 tables in the lower left, the left column gives the number of points above the average LOC and the right column gives the number of points below the average LOC; the upper row gives the points for the better condition and the lower row gives the points for the poorer condition.

DISCUSSION

The principal finding was that the LOC varied with the length and variability of the foreperiod. On the assumption that the LOC represents a limitation on performance that is invariant under changes in the criterion for emitting a response, then temporal uncertainty causes changes in the speed of information processing.

Temporal uncertainty apparently also produces a change in the response criterion. Variation in response criterion was most evident when the foreperiod was randomly 750 or 1,250 msec. Stimuli presented after the longer of the two foreperiods could more easily be anticipated, and the RTs were faster on those trials, but the longer foreperiod also produced a higher error rate. When the LOC was used to portray the reciprocity between speed and accuracy, there was little or no difference between these conditions. Similarly, other differences in RT distributions were associated with reduced differences in LOCs. These results are consistent with the hypothesis that contributions of the response criterion are removed from the LOC although the evidence for this hypothesis would have been much more dramatic if the LOCs had remained invariant across all of the foreperiod conditions.

The surprising finding was that performance was more efficient with the 975-msec. foreperiod than with foreperiods that were constant (1,000 msec.) or slightly longer (1,025 msec.). This finding seems paradoxial because *S* should have been less well prepared for the early signal than for the constant or the later signals. Why should performance be more efficient when the stimulus is presented before *S*'s expectancy has reached a maximum? The result is contrary to both the Anticipation model and the Preparatory Set model. One explanation can be provided by the Diluted Sample model. By this model, uncertainty about the point of stimulus onset reduces the extent to which *S*'s sample of sensory data coincides with the actual stimulus presentation. Thus, when expectancy is lower at earlier points in the trial, the less irrelevant preliminary data are likely to be included in the observational sample of sensory output. This explanation is consistent with the greater efficiency under the shorter foreperiod in Group U (at 975 msec.) and also with the poorer performance under the most variable foreperiods (750 and 1,250 msec. in Group N). This explanation is, however, post hoc and begs additional confirmation.

Interestingly, the effects of temporal uncertainty were reflected mainly in the slopes of the LOCs, i.e., mainly at the longer RTs and lower error rates. The estimated difference between the best and worst conditions (U 975 and N 1,250) was 13 msec. at the point where discrimination began to rise above chance and was approximately 95 msec. where $d' = 4.0$ (corresponding to an error rate of about .02). But the greater difference in RTs at low error rates does not point to the advisability of conducting experiments with low error rates because estimates of the error rate become less reliable as the error rate is reduced. The point is that different processes seem to underlie the limitation measured by the LOC and the speed of performance as measured by mean RT. Presumably, the contribution of a response criterion is removed from the LOC.

REFERENCES

Bertelson, P. The time course of preparation. *Quarterly Journal of Experimental Psychology,* 1967, **18**, 272–279.

Green, D. M., & Luce, R. D. Detection of auditory signals presented at random times: III. *Perception & Psychophysics,* 1971, **9**, 257–268.

Karlin, L. Reaction time as a function of foreperiod duration and variability. *Journal of Experimental Psychology,* 1959, **58**, 185–191.

Lappin, J. S., & Disch, K. The latency operating characteristic: I. Effects of stimulus probability on choice reaction time. *Journal of Experimental Psychology,* 1972, **92**, 419–427. (a)

Lappin, J. S., & Disch, K. The latency operating characteristic: II. Effects of visual stimulus intensity on choice reaction time. *Journal of Experimental Psychology,* 1972, **93**, 267–372. (b)

Nickerson, R. S. Response time to the second of two successive signals as a function of absolute and relative duration of intersignal interval. *Perceptual and Motor Skills,* 1965, **21**, 3–10.

Posner, M. I., & Boies, S. J. Components of attention. *Psychological Review,* 1971, **78**, 391–408.

Siegel, S. *Nonparametric statistics for the behavioral sciences.* New York: McGraw-Hill, 1956.

Snodgrass, J. C. Foreperiod effects in simple reaction time: Anticipation or expectancy? *Journal of Experimental Psychology Monograph,* 1969, **79** (3, Pt. 2).

Thomas, E. A. C. Reaction-time studies: The anticipation and interaction of responses. *British Journal of Mathematical and Statistical Psychology,* 1967, **20**, 1–29.

Woodrow, H. The measurement of attention. *Psychological Monographs,* 1914, **17** (5, Whole No. 76).

Yellott, J. I. Correction for fast guessing and the speed-accuracy tradeoff in choice reaction time. *Journal of Mathematical Psychology,* 1971, **8**, 159–199.

(Received June 29, 1972)

Journal of Experimental Psychology
1973, Vol. 98, No. 2, 286–290

EFFECTIVENESS OF SUPPLIED MEDIATORS IN RELATION TO PRESENTATION MODALITY AND RETRIEVAL CUE[1]

TANNIS Y. ARBUCKLE[2] AND LOUISE AZNAVOUR[3]

Sir George Williams University, Montreal, Quebec, Canada

Effectiveness of supplied mediators was studied by having mediators supplied auditorily (A supplied) or visually (V supplied), generated by *S* (generated), or not mentioned in the instructions (control). Two hundred and forty *Ss* learned two lists, each containing six abstract noun – concrete noun (A–C) pairs with the A noun as test cue (A–C/A) and six concrete noun – abstract noun (C–A) pairs with the C noun as test cue (C–A/C), under one of eight combinations of four mediational conditions and two presentation rates. On two measures of learning there was no difference between the generated and A-supplied conditions or between the V-supplied and control conditions, but the former two conditions were superior to the control condition on both measures and to the V-supplied condition on one measure. Learning was faster with the slower presentation rate, the second list, and the C–A/C pairs, with the C–A/C superiority being accentuated with supplied mediators. Results were related to availability and accessibility of imaginal representations of supplied mediators.

Recent studies of paired-associate (PA) learning which compared mediators supplied by *E* with those generated by *S* have found significantly greater recall with generated mediators (Bobrow & Bower, 1969; Bower & Winzenz, 1970; Pelton, 1969; Schwartz, 1971). These results have been obtained with a variety of supplied mediators: short sentences constructed by *E* in which the two words were embedded (Bobrow & Bower, 1969; Bower & Winzenz, 1970), sentences produced by a yoked control *S* (Pelton, 1969), and single words which were known to be primary mediators for the pairs (Schwartz, 1971). An exception to the general finding of significant superiority of generated mediators is a study by Lieberman, Walters, and Cox (1968) which found no difference between the two conditions. The Lieberman et al. study differed from all of the experiments cited except Pelton's in that the mediators were supplied auditorily; it differed from Pelton's study in that the supplied mediators were popular rather than idiosyncratic.

When groups supplied with mediators have been compared to uninstructed control *Ss*, significant facilitation of the learning of adult *Ss* has been found with letter–word pairs and single-word mediators (Schwartz, 1971; Wood & Bolt, 1968), but not with syllable–word pairs and single-word mediators (Dallett, 1964) or with word–word pairs and sentence mediators (Duffy & Montague, 1971). In all of these experiments the pairs and mediators were presented visually.

The obvious explanation of the inefficiency of supplied mediators is that adult *Ss* normally generate at least some mediators even without special instructions and that mediators supplied by *E* are simply not as appropriate for an individual *S* as are those he generates himself. However this explanation is at best incomplete, given Schwartz's (1971) finding that generated mediators were more effective than supplied mediators even when the comparison was restricted to those cases where the generated and supplied mediators were identical. Schwartz reported that the superior recall of responses found with generated mediators was attributable primarily to the greater retrievability of the generated as compared with the supplied mediator at time of testing. She suggested that this greater retrievability could be due

---

[1] This research was supported by Grant APA 209 from the National Research Council of Canada.

[2] Requests for reprints should be sent to Tannis Y. Arbuckle, Department of Psychology, Sir George Williams University, Montreal 107, Quebec, Canada.

[3] Now at the University of Montreal, Montreal, Quebec, Canada.

to (*a*) confusion between the supplied mediator and the response, (*b*) greater similarity of study and test conditions with generated mediators, (*c*) more meaningful relations between pairs and mediators when *S* generates them himself (essentially the "comprehension" hypothesis of Bobrow and Bower, 1969), or (*d*) motivational benefits accruing to *S* who discovers mediators by himself.

The present experiment was an attempt to ascertain more exactly the conditions determining the effectiveness of supplied mediators. To this end, learning was examined for two successive lists and two rates of presentation, with mediators being supplied either auditorily or visually and with strong and weak retrieval cues. The definition of strong and weak retrieval cues was based on Lockhart's (1969) finding that concrete (C) nouns are more effective cues for retrieval than are abstract (A) nouns.

METHOD

The *S*s were 240 undergraduate students who volunteered to serve in the experiment.

Four mediational conditions—with mediators being (*a*) supplied visually (V supplied), (b) supplied auditorily (A supplied), (*c*) generated by *S* in response to instructions to form words, images, or phrases to combine the word pair (generated), and (*d*) not mentioned in *S*'s instructions (control)—were combined factorially with two presentation rates, 4 and 6 sec., to yield eight experimental groups with 30 *S*s being assigned in rotation to each group. All *S*s, tested individually, learned two 12-pair lists to a criterion of one perfect recitation using the study–test method of PA learning. Each list served as first list learned for half of the *S*s in each group. All pairs consisted of a C and an A noun selected from Paivio, Yuille, and Madigan (1968), with six pairs in each list being randomly assigned to the C–A order with the C noun as test cue (C–A/C) and six to the A–C order with the A noun as test cue (A–C/A). The same two lists were used by Arbuckle (1971) and are further described there. The supplied mediators were derived from those reported by *S*s in the previous Arbuckle study by abstracting the most frequently occurring theme from the mediators reported for each pair and expressing this theme in a sentence. These sentence mediators differed from sentences supplied by other *E*s (Bobrow & Bower, 1969; Bower & Winzenz, 1970; Duffy & Montague, 1971) in that they were more variable in length (range: 5–11 words), greater in complexity (e.g., 19 out of 24 contained at least one additional noun), and greater in degree of transformation of pair members (e.g., order of occurrence of pair members was reversed in the sentences for 5 A–C and 6 C–A pairs; several nouns were changed in form: to verbs [1C, 3A], to adjectives [4 A], to related C nouns [2 A], to plurals [6 C], and to possessives [1 C]). The pairs and left-hand members were photographed on 2 × 2 negative slides, the mediators in the V-supplied condition being placed below the pair. The mediators in the A-supplied condition were presented by a Uher 8000 tape recorder, the Diapilot of which was used to pace the projection (by a Kodak Carousel projector) of the visual materials under all conditions and to ensure approximate simultaneity of onset of pair and mediator in the A-supplied condition. Approximate simultaneity of offset was achieved by speaking the mediators over 4 or 6 sec. Four different orders of pairs were constructed for each list, with starting point in the sequence of orders being systematically varied across *S*s within groups and each *S* receiving the orders in sequence until criterion was reached. Response members, recalled orally, had to be exactly reproduced to be counted correct. Immediately after reaching criterion on the second list in the control condition and on each list in the other three conditions, *S*s were questioned on use of mediators and were asked to reproduce the mediators that were supplied or generated.

RESULTS

Analysis of trials to criterion as a function of rate, list, and mediational condition showed that learning was significantly faster with the 6-sec. rate, F (1, 232) = 22.29, $p < .01$, and the second list, F (1, 232) = 51.98, $p < .01$; that it differed significantly across mediational conditions, F (3, 232) = 3.66, $p < .05$; and that there were no interactions. Post hoc Newman-Keuls tests showed that learning was significantly ($p < .05$) faster under generated and A-supplied conditions than under the control condition. Means and standard deviations of trials to criterion are given in Table 1.

Analysis of number of responses correctly recalled on Trial 1 as a function of rate, list, pair type, and mediational condition showed that recall was significantly greater with the 6-sec. rate, F (1, 232) = 14.63, $p < .01$; the second list, F (1, 232) = 51.92, $p < .01$; and the C–A/C pairs, F (1, 232) = 8.55, $p < .01$; and that it varied significantly across mediational condition, F (3, 232) = 6.81, $p < .01$. Post hoc Newman-Keuls tests showed that there

TABLE 1

MEANS AND STANDARD DEVIATIONS OF TRIALS TO CRITERION AS A FUNCTION OF PRESENTATION RATE, LIST, AND MEDIATIONAL CONDITION

Rate	List	Mediational condition							
		Control		Visually supplied		Auditorily supplied		Generated by *S*	
		M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
4 sec.	First	5.3	2.1	5.3	2.0	4.6	2.9	4.4	2.6
	Second	4.6	2.0	4.2	2.6	3.7	2.5	4.0	2.2
6 sec.	First	4.5	2.4	4.0	2.5	4.0	2.6	3.8	2.2
	Second	3.6	2.9	3.5	2.2	3.1	2.4	2.9	2.6
M		4.50_{a}		$4.25_{a,b}$		3.85_{b}		3.77_{b}	

Note. Cells having different subscripts are significantly different at the .05 level.

was no difference between the generated and A-supplied conditions or between the V-supplied and control conditions, but that the former two conditions were significantly superior to the latter two conditions. There were two significant interactions. As may be seen from Table 2 the Rate × List interaction, F (1, 232) = 4.04, $p < .05$, reflected a greater first-to-second list improvement with the 6-sec. rate, while the Pair Type × Mediational Condition interaction, F (3, 232) = 2.87, $p < .05$, reflected an accentuation of the superiority of C–A/C over A–C/A pairs with supplied mediators. Post hoc Scheffé tests on this interaction using Cicchetti's (1972) technique showed that with C–A/C pairs the A-supplied condition was significantly ($p < .05$) superior to the control and V-supplied conditions and not different from the generated condition, while the V-supplied condition did not differ from either the control or generated conditions. With A–C/A pairs, neither supplied condition differed from the control condition, and both were significantly ($p < .05$) inferior to the generated condition.

Analysis of mediators reported generated on the second list as a function of mediational condition and rate showed that more pairs were reported mediated under the generated condition (83%) than under the

TABLE 2

CORRECT RECALL ON TRIAL 1 AS A FUNCTION OF PAIR TYPE, MEDIATIONAL CONDITION, LIST, AND PRESENTATION RATE

Pair type	Mediational condition	List and rate				*M*
		First		Second		
		4 sec.	6 sec.	4 sec.	6 sec.	
C–A/C	Control	2.07	1.87	1.97	2.63	2.13
	Visually supplied	1.83	2.57	2.23	3.00	2.41
	Auditorily supplied	2.43	2.50	3.10	3.63	2.92
	Generated by *S*	2.53	2.60	2.73	3.50	2.84
A–C/A	Control	1.27	1.97	2.13	2.83	2.05
	Visually supplied	1.40	2.00	2.07	2.93	2.10
	Auditorily supplied	1.80	2.30	2.37	3.00	2.36
	Generated by *S*	2.53	2.83	2.80	3.37	2.88
M		1.98	2.33	2.42	3.13	2.46

Note. Pair types: C–A/C = Concrete (C) noun paired with abstract (A) noun with C noun as test cue; A–C/A = A–C ordered pair with A noun as test cue.

control condition (53%), F (1, 116) = 35.8, $p < .01$, while analysis of reported mediators in the generated condition as a function of list, rate, and pair type showed more A–C pairs (85%) than C–A pairs (79%) were reported mediated, F (1, 58) = 9.0, $p < .01$. No other effects in either analysis were significant.

In the V-supplied and A-supplied conditions *S*s generally reported having used the supplied mediators for most of the pairs. When the reliability of this report was checked by having *S*s recall or paraphrase the mediators, only eight A-supplied and nine V-supplied *S*s failed to reproduce satisfactorily at least 20 out of 24 mediators. All but four of these failures were with the 4-sec. rate.

DISCUSSION

The main interest of this experiment lies in the finding that supplied mediators, when presented auditorily and used with C–A/C pairs, were as effective as *S*-generated mediators.

The modality effect is consistent with previous findings of Lieberman et al. (1968) that *S*s given mediators auditorily learned as well as *S*s who generated mediators and of Duffy and Montague (1971) that *S*s given mediators visually learned no faster than uninstructed controls. The fact that mediators supplied auditorily facilitated learning relative to the uninstructed controls, while the same mediators presented visually did not do so, could be explained by Schwartz's (1971) suggestion that *S*s confuse the pair with the mediator. Specifically, presenting the pair visually and the mediator auditorily should serve to increase the discriminability of the pair from the mediator, thereby reducing potential confusion and increasing correct recall.

None of Schwartz's (1971) four suggestions, however, seems to explain why supplied mediators were more effective with C–A/C pairs than with A–C/A pairs. It is of course possible that this finding was an artifact of a "better" set of mediators having been supplied for the C–A/C pairs, although since mediators were constructed in the same manner for all pairs and pairs were randomly assigned to conditions, this possibility should have been minimal. In subsequent work we checked mediator recall for these two sets of pairs using an incidental learning task and a new group of *S*s. Overall mediator recall was similar for the two sets of pairs (47% of the mediators were recalled for C–A/C pairs and 44% for the A–C/A pairs), but within each set mediator recall was much more likely when cued by the C member than when cued by the A member. These data are consistent with the argument that the greater effectiveness of the supplied mediators with the C–A/C pairs in the present experiment was not an artifact of the particular mediators used, but occurred because the C nouns were better retrieval cues for the mediators than were the A nouns.

One way of accounting for the varying effectiveness of supplied mediators across experimental conditions is to assume, on the basis of recent statements about the role of imagery in language (e.g., Paivio, 1971a), that the effectiveness of a supplied verbal mediator depends on the extent to which *S*s can create and retrieve an imaginal representation of it. The findings of Atwood (1971) suggest that visual presentation of a verbal mediator would interfere with image creation while auditory presentation would not. If so, then in the A-supplied condition *S*s would be more likely to have an image as well as a verbal mediator available for each pair, a situation which should, according to the dual-coding hypothesis (Paivio, 1971b), result in greater recall with A-supplied mediators. Furthermore, if the supplied mediators were being transformed into images, the greater effectiveness of the C noun as a retrieval cue for the mediator and the responses would be explicable, since C nouns would be presumed to have greater redintegrative capacity for images than would A nouns.

What this explanation does not account for is why there was not a similar advantage of C–A/C pairs over A–C/A pairs in the generated condition since presumably these *S*s were also using imaginal as well as verbal mediators. One possibility is that since these *S*s reported mediating more A–C/A than C–A/C pairs, this "extra effort" may have compensated for a retrieval cue effect.

REFERENCES

ARBUCKLE, T. Y. Mediational instructions, stage of practice, presentation rate, and retrieval cue in paired-associate learning. *Journal of Experimental Psychology*, 1971, **88**, 396–402.

ATWOOD, G. An experimental study of visual imagination and memory. *Cognitive Psychology*, 1971, **2**, 290–299.

Bobrow, S. A., & Bower, G. H. Comprehension and recall of sentences. *Journal of Experimental Psychology*, 1969, **80**, 455–461.

Bower, G. H., & Winzenz, D. Comparison of associative learning strategies. *Psychonomic Science*, 1970, **20**, 119–120.

Cicchetti, D. V. Extension of multiple-range tests to interaction tables in the analysis of variance: A rapid approximate solution. *Psychological Bulletin*, 1972, **77**, 405–408.

Dallett, K. M. Implicit mediators in paired-associate learning. *Journal of Verbal Learning and Verbal Behavior*, 1964, **3**, 209–214.

Duffy, T. M., & Montague, W. E. Sentence mnemonics and noun pair learning. *Journal of Verbal Learning and Verbal Behavior*, 1971, **10**, 157–162.

Lieberman, L. R., Walters, W. M., & Cox, W. M. Experimenter-made vs. subject-made mnemonics in paired-associate learning. *Psychological Reports*, 1968, **22**, 227–231.

Lockhart, R. S. Retrieval asymmetry in the recall of adjectives and nouns. *Journal of Experimental Psychology*, 1969, **79**, 12–17.

Paivio, A. Imagery and language. In S. J. Segal (Ed.), *Imagery: Current cognitive approaches*. New York: Academic Press, 1971. (a)

Paivio, A. *Imagery and verbal processes.* New York: Holt, Rinehart & Winston, 1971. (b)

Paivio, A., Yuille, J. C., & Madigan, S. Concreteness, imagery, and meaningfulness values for 925 nouns. *Journal of Experimental Psychology Monograph Supplement*, 1968, **76** (1, Pt. 2).

Pelton, L. H. Mediational construction vs. mediational perception in paired-associate learning. *Psychonomic Science*, 1969, **17**, 220–221.

Schwartz, M. Subject-generated versus experimenter-supplied mediators in paired-associate learning. *Journal of Experimental Psychology*, 1971, **87**, 389–395.

Wood, G., & Bolt, M. Mediation and mediation time in paired-associate learning. *Journal of Experimental Psychology*, 1968, **78**, 15–20.

(Received July 3, 1972)

Journal of Experimental Psychology
1973, Vol. 98, No. 2, 291–296

SOME STIMULUS VARIABLES AFFECTING SOLUTION SHIFT PERFORMANCE[1]

L. E. BOURNE, JR.[2] AND SHIRLEY MILLER

Institute for the Study of Intellectual Behavior, University of Colorado

Conceptual problems involving an unannounced shift in solution have typically taken three forms: an intradimensional shift with reversal of stimulus–response categories, an extradimensional shift to a new relevant dimension, and an intradimensional shift to new values on the preshift relevant dimension. Two variables, the novelty of new stimulus values introduced at the time of the shift and the similarity of postshift to preshift stimulus values, which had not been systematically examined, are shown in this experiment to affect the order of shift difficulty. Implications for theories of shift behavior and a possible integration of available evidence are presented.

Concept problems which involve an unannounced shift in solution have provided a method for evaluating a variety of theoretical ideas (Johnson, 1966; Kendler & Kendler, 1969; Polson & Dunham, 1971). Several forms of shift have been explored. The intradimensional reversal shift (IDR), a negative transfer paradigm, imposes a change in all initially learned stimulus-category assignments. If *S* has learned to place red stimuli in Category A and green stimuli in Category B, an IDR shift requires him to place red stimuli in Category B and green stimuli in Category A. The extradimensional shift (ED) imposes a change of relevant stimulus dimensions. Preshift solution might require *S* to place squares in Category A and triangles in Category B. After the shift, reds are assigned to Category A and greens to Category B. The fact that human adult *S*s typically perform better on an IDR shift in contrast to an ED shift has been taken by Kendler and Kendler (1969) to indicate the necessity of a mediational interpretation of adult conceptual behavior. During preshift learning, *S* develops the appropriate dimensional mediator for post-IDR performance but not for post-ED performance.

More recently, Isaacs and Duncan (1962) and Johnson (1966) have examined and compared an intradimensional nonreversal shift (IDN), a positive transfer paradigm, with IDR and ED shifts. In an IDN shift, the same dimension is relevant before and after the shift, but different values are used to represent it. Prior to the shift, *S* may learn to place yellow stimuli in Category A and blue stimuli in Category B. After an IDN shift, he must place reds in Category A and greens in Category B. Like the IDR shift, the IDN shift should be, and is, significantly easier than the ED shift for adult *S*s (Isaacs & Duncan, 1962; Johnson, 1966), presumably because the same dimensional mediator holds before and after. Moreover, the IDN shift is significantly easier than the IDR shift. Isaacs and Duncan account for this fact with the argument that particular stimulus-category assignments learned during the preshift phase do not have to be extinguished after an IDN, whereas they do after an IDR shift. With an IDN shift, the newly required associations can begin to form immediately, for the old stimulus values have been removed altogether from the task. This result, and another which shows the differential effects of number of irrelevant dimensions on postshift performance, has led Johnson to an

---

[1] This research was undertaken within the Institute for the Study of Intellectual Behavior and is Publication No. 29 of the Institute. The work was supported by Research Grant MH-14314 and Research Scientist Award 1-K5-MH-37497, both from the National Institute of Mental Health, and Research Grant GB-340-77X from the National Science Foundation. The authors wish to thank C. C. Hiew and Linda Crews for collecting and analyzing the data and Paul Neumann for critical comments on an earlier version of the manuscript.

[2] Requests for reprints should be sent to Lyle E. Bourne, Jr., Institute for the Study of Intellectual Behavior, University of Colorado, Boulder, Colorado 80302.

alternative hypothesis-testing interpretation of solution shift experiments.

Comparisons of the sort reported by Isaacs and Duncan (1962) and Johnson (1966) suffer from two methodological problems which obscure their interpretation. First of all, in Johnson's study, only the stimuli of the IDN condition were changed at the time of the shift. When the shift involved an IDR or an ED shift, the stimulus population was exactly the same throughout the experiment. A noticeable change in the stimuli could have unknown effects attributable to novelty or attentional factors (Guy, Van Fleet, & Bourne, 1966). The proper comparison among shifts ought to include comparable stimulus changes in all cases, as in Isaacs and Duncan. Second, when stimulus changes have been made, they have been of arbitrary magnitude. Whether *S* forms associations between entirely new values on the relevant dimensions after an IDN shift or whether these associations are affected by some form and degree of generalization is not clear. In order to decide that question, the similarity between pre- and postshift stimulus values has to be investigated.

In the present experiment, degree of similarity of pre- and postshift values on one stimulus dimension, size, was examined in IDR, IDN, and ED shifts. For the IDR shift, size was relevant before and after the shift. If the smaller of the two values was positive before the shift, the larger of the two was positive after the shift. Likewise, size was relevant before and after the IDN shift. If the smaller value was relevant beforehand, the smaller one was also relevant after the shift. For ED, size was relevant in the postshift phase, but a different dimension (color) was relevant in the preshift phase. Across shift conditions, the values of size during pre- and postshift stages bore one of four similarity relationships. If there is some generalization from preshift to postshift stimulus values, then the following effects might be expected. In the ED case, the effects should be negligible. Assuming that the postshift cues are discriminable, performance should be about the same regardless of the similarity relationship of cues, *unless there is a novelty effect associated with the stimulus change* which increases with the degree of change. In the IDN case, postshift performance should be better the more similar pre- and postshift cues. The *S* has learned a pair of associations in the preshift phase. The same stimulus–response relationship holds in the postshift phase. Therefore, the greater the similarity between the pre- and the postshift cues, the greater the generalization from preshift associations to postshift associations. Essentially, the same expectation holds in reverse for IDR shifts. That is, an IDR shift should be easier to execute as the difference between pre- and postshift values on the relevant dimension increases. The more similar the pre- and postshift cues are, the greater the generalization and, therefore, the greater the preshift associative strength that must be overcome before a postshift solution can be acquired. In comparing IDN and IDR postshift performance, we expect the typical relationship to hold, IDN easier than (<) IDR, when pre- and postshift cues are highly similar. Indeed, we expect IDN < IDR < ED. As similarity decreases, however, the difference between IDN and IDR should become smaller and possibly be eliminated.

One additional factor was introduced into this experiment, the difficulty of the discrimination (called here discriminability) required in the postshift phase. Eight discriminable values on the size dimension, A, B, C, D, E, F, G, and H, running from largest to smallest, were used. Half of the *Ss* in the postshift phase were required to perform on the basis of an A vs. H (AH) discrimination and the remaining *Ss* solved on the basis of D vs. E (DE). The expected effect is clear. The AH problem should be significantly easier than the DE problem (Archer, 1962). But, discriminability might interact with pre- and postshift stimulus similarity and type of shift. Once again, the ED shift should be unaffected except for a general elevation in trials and errors with the more difficult discrimination. For the IDN and IDR shifts, stimulus generalization effects may be potentiated with the more difficult postshift problem, if only because there is more room for an effect to manifest itself.

The purpose of this experiment, then, is to explore the effects of pre- and postshift stimulus similarity, type of shift, and difficulty of discrimination in the standard solution shift – concept identification paradigm. The results are expected to clarify not only some theoretical questions about the importance of stimulus generalization in solution shift performance, but also to provide information on certain previously unexplored methodological variables.

METHOD

Subjects and design. One hundred ninety-two college students, 96 males and 96 females averaging 19 yr. of age, were paid $1.50 to participate as *S*s. Equal numbers of male and female *S*s, all naive to concept-learning experiments, were assigned at random to the 24 experimental conditions.

The basic design was a 4 × 3 × 2 factorial. There were four pairs of preshift relevant attributes, differing in intrapair discriminability (AH, BG, CF, DE). In the AH condition, one half of the *S*s learned problems in which the stimulus size A was positive and size H was negative; the other half of the *S*s learned the reverse contingency. Conditions BG, CF, and DE were similarly designed. Three shifts were used: IDN, IDR, and ED. In Shift IDR, the positive and negative (size) cues were reversed after the shift; in IDN, positive and negative (size) cues stood in the same relationship before and after the shift; and in ED, cues along a preshift irrelevant dimension (size) were relevant after the shift. Color was the preshift relevant dimension for ED groups. There were two conditions of postshift intrapair discriminability (AH and DE).

The *S*s sat before a screen onto which slides were projected, one at a time, by a Kodak Carousel projector. The *S*s responded manually, by pressing one of two buttons labeled with a + or −. Each slide remained in view until *S* made a response. Following each response a feedback light appeared for a 2-sec. interval over the correct response button. A 5-sec. intertrial interval preceded presentation of the next stimulus. Every trial, then, consisted of a stimulus, *S*'s category response, and a feedback signal.

Stimuli. Stimuli were geometric figures varying on three two-valued dimensions: color (red/blue), shape (square/triangle), and number (one/two figures); and one eight-valued dimension, size. The sizes, labeled A, B, C, D, E, F, G, and H, differed in base length intervals of 1.8 cm. when projected. Figures were constructed on cards, photographed, and mounted as slides. Stimulus sequences included equal numbers of randomly ordered positive and negative instances, and equal numbers of occurrences of values on the irrelevant dimensions.

Task and procedure. General instructions and a description of the task were given to all *S*s before the experimental problems were administered. The *S*s were told that the task was to learn a rule for assigning stimuli to two categories. The kind of responses required was explained. No indication was given that a shift in solution would occur. After 50 preshift trials, any *S* who had not begun a criterion run of 12 consecutive correct responses was informed that a single attribute was positive and was reminded of the four dimensions of the stimuli; the procedure then continued. If *S* had not reached criterion after 120 trials, he was eliminated from the experiment. Five *S*s were eliminated as nonsolvers; they were distributed without apparent bias across experimental conditions. When *S* had reached preshift criterion, the shift was instituted without interruption.

RESULTS

Preliminary tests revealed that raw-score trials and errors to solution did not conform to the *F*-test assumptions of equality of variance and normality. When the scores were transformed exponentially, abnormalities of the data were eliminated. Analyses of variance on the transformed

TABLE 1

MEAN TRIALS TO SOLUTION

Shift	Preshift stimulus values			
	AH	BG	CF	DE
Preshift phase				
ED	11.5	12.3	8.9	9.4
IDR	19.7	10.3	20.1	50.3
IDN	13.3	13.5	15.8	44.7
Postshift phase (Postshift stimulus value = AH)				
ED	23.0	16.1	15.8	19.9
IDR	13.8	10.1	4.4	5.6
IDN	.3	1.0	1.6	3.0
Postshift phase (Postshift stimulus value = DE)				
ED	62.0	57.8	50.8	59.3
IDR	61.4	68.3	50.8	13.9
IDN	59.9	69.1	11.9	3.0

Note. Shifts: ED = extradimensional; IDR = intradimensional reversal; and IDN = intradimensional nonreversal. Stimulus values: A, B, C, D, E, F, G, and H represent eight discriminable values, and letter pairs indicate differing intrapair discriminability.

and raw scores on trials and errors showed identical effects. Therefore, only analyses of raw trials are reported.

Preshift performance. Mean trials in various Shift × Preshift Discriminability conditions are presented in Table 1. Statistically significant effects attributable to both main variables were observed for stimulus discriminability, $F\ (3,\ 144) = 17.10$, $p < .01$, and for shift condition, $F\ (2,\ 144) = 11.95$, $p < .01$. More informative is the Shift × Discriminability interaction, $F\ (6,\ 144) = 5.33$, $p < .01$. Discriminability affects performance only in IDR and IDN conditions, and not in ED. This was, of course, expected because the dimension whose values varied in discriminability was relevant in the preshift phase only for IDR and IDN conditions.

Individual F tests were performed comparing IDR and IDN groups at each discriminability level. None was statistically significant. Similar comparisons were made between IDN *S*s, who solved the preshift phase on the basis of the color dimension, and the combined IDN and IDR groups, who solved on the basis of size. The IDR and IDN groups performed reliably worse than the ED group only at discriminability level DE, $F\ (1,\ 144) = 9.02$, $p < .01$. A trend analysis was performed to determine whether changes in size discriminability had an effect on preshift ED performance (size irrelevant). Although there was a general decrease in mean trials to solution with decreased discriminability, there were no significant components of trend.

Postshift performance. The mean trials to postshift criterion are given in Table 1. The main effects of shift and preshift discriminability were vastly (and statistically) different, depending on postshift stimulus discriminability. Therefore, separate analyses of variance were conducted for the two postshift conditions. For postshift condition DE, type of solution shift, $F\ (2,\ 84) = 11.19$, $p < .01$, and preshift discriminability, $F\ (3,\ 84) = 11.05$, $p < .01$, differed reliably overall. Moreover, there was a Shift × Preshift Discriminability interaction, $F\ (6,\ 84) = 3.83$, $p < .01$, with IDR and IDN shifts absorbing most of the discriminability effect. A separate analysis on only the IDR and IDN scores revealed a small, though significant interaction, $F\ (1,\ 84) = 2.83$, $p < .05$. Performance in these conditions, indeed in all three shifts, converges with decreasing pre- and postshift stimulus similarity.

The only reliable effect in postshift condition AH was shift, $F\ (2,\ 84) = 8.64$, $p < .01$. Further analyses of the individual shifts revealed only one trend, a linear reduction in trials for IDR with decreasing preshift discriminability, $F\ (1,\ 28) = 4.32$, $p < .05$. A trend analysis of the IDN points failed to reach significance even though IDN *S*s had increasing difficulty as the postshift stimulus values became less similar to those of the original pair. In a two-way analysis of IDN and IDR, the Shift × Preshift Discriminability interaction was marginally significant, $F\ (3,\ 56) = 2.12$, $p < .10$.

As a final analysis, the IDR, IDN, and ED groups were compared at those points where no change in stimulus values occurred with the shift. Thus, only those groups with either AH or DE values on both tasks were included. Effects of shift, $F\ (2,\ 42) = 21.70$, $p < .01$; postshift stimulus discriminability, $F\ (1,\ 42) = 6.02$, $p < .01$; and their interaction, $F\ (2,\ 42) = 5.02$, $p < .05$, were all reliable. Although the ordering of shift difficulty (IDN < IDR < ED) was the same when either the AH or DE values were involved in both tasks, all shifts were more difficult when the DE contingencies were learned.

DISCUSSION

When postshift relevant cue values were easily discriminated (AH), there were no changes with preshift stimulus discriminability in the usual ordering of shift difficulty (e.g., Isaacs & Duncan, 1962). In all cases IDN < IDR < ED. While IDN postshift performance was essentially unaffected, trials to achieve criterion decreased after an IDR shift and increased after an IDN shift as the similarity of pre- and postshift cues decreased. These latter trends are consistent with simple stimulus generalization effects. For IDN, *S* must respond in the same way after the shift to decreasingly similar stimuli, resulting in some minor deterioration of performance. For IDR, the change of cues is facilitative, because of a resultant *decrease* in the tendency to make the same response with decreasing similarity.

When postshift relevant cue values were difficult to discriminate (DE), the order of IDR, ED, and IDN shifts was altered radically by changing the discriminability of cues during the preshift phase. Again, ED performance was essentially unaffected. Performance on both IDR and IDN shifts was facilitated by increasing the similarity of pre- and postshift cues, even though these two shifts tend to maintain their relative positions of difficulty. All shifts are essentially equivalent in difficulty when the difference between pre- and postshift relevant cue values is large. As that difference decreases, the usual shift ordering, IDN < IDR < ED, manifests itself.

The lack of any effect of preshift stimulus discriminability on ED performance is, once again, understandable when one considers that the variable dimension was irrelevant in the preshift phase of the experiment. On a stimulus generalization argument, we would have expected IDN performance to get worse and IDR performance to get better with decreasing similarity between pre- and postshift cues. In fact, the absolute difference between shifts does decrease, but the major effect of this variable is to make performance on IDR and IDN shifts worse. Combined with the relatively minor effects observed with the easy postshift problem (AH), these results suggest that stimulus generalization makes, at best, a small contribution at shift tasks.

The major effect of preshift discriminability appears to be in an alerting function with respect to the relevant dimension. When there is a large change, from AH to DE, in the discriminability of values taken from the relevant dimension at the time of the shift, *S*s (*a*) are alerted to a change in the nature of the problem, and (*b*) may fail to recognize any variation at all in the formerly relevant dimension. At the shift, *S*s seek a new solution, basing their hypotheses on some dimension other than the one relevant in the preshift phase. This works adversely on performance in both the IDR and IDN conditions. As the amount of change is reduced, moving through preshift values BG and CF, *S* becomes increasingly more sensitive to the possibility of a new solution based on the preshift relevant dimension.

If it were simply a matter of a large pre- to postshift stimulus change serving to alert *S* that the solution has shifted, then postshift performance in the DE condition should be essentially identical to performance in the AH condition because the change is equal, although in the opposite direction. This is clearly not the case. Postshift performance is retarded as a function of the collapse of the preshift difference between stimulus values, i.e., in the XX to DE conditions, but remains fairly stable as a function of an increase in the preshift difference, i.e., in XX to AH conditions. This suggests that those *S*s who learn a DE discrimination during the preshift phase have become more highly sensitized to differences in size than have *S*s in the preshift AH condition. Pre- and postshift collapse of the difference in size produces inferior performance, probably because size is no longer perceived as a variable dimension by at least some *S*s.

These findings imply that both pre- and postshift stimulus discriminability and the relationship of stimulus values in the pre- and postshift phases can be significant factors affecting the relative difficulty of solution shifts in conceptual problems. When the postshift task involves a relatively easy discrimination, the results of Isaacs and Duncan (1962) are replicable regardless of the similarity between pre- and postshift cues. The IDR is universally easier than the ED, possibly because no change of mediational response is required in the IDR case. The IDN is easier than the IDR, possibly because not only the mediator remains the same, but the specific stimulus–response assignments required in a postshift phase benefit from modest stimulus generalization effects.

The relationship between pre- and postshift cues becomes critical when the postshift task requires a difficult discrimination. With relatively little difference in cues, the results of Isaacs and Duncan (1962) are once again replicated. As the difference increases, a point is reached (in this case, CF) at which IDN < IDR = ED, the ordering obtained by Johnson (1966). When the difference between pre- and postshift cues is large, shift differences are eliminated, suggesting that, at the time of the shift, *S*s reset to zero and begin a new search for a solution to the problem. We conclude from these results that between-phase stimulus similarity is a significant variable controlling problem solving performance and relative shift difficulty, especially when the required discriminations are hard on an absolute scale. The results are not inconsistent with either mediational (Kendler & Kendler, 1969) or hypothesis-testing (Erickson, 1971; Johnson, 1966) theory, but indicate needed directions of elaboration in both.

REFERENCES

Archer, E. J. Concept identification as a function of obviousness of relevant and irrelevant information. *Journal of Experimental Psychology*, 1962, **63**, 616–620.

Erickson, J. R. Problem shifts and hypothesis behavior in concept identification. *American Journal of Psychology*, 1971, **84**, 100–111.

Guy, D. E., Van Fleet, F., & Bourne, L. E., Jr. Effects of adding a stimulus dimension prior to a nonreversal shift. *Journal of Experimental Psychology*, 1966, **72**, 161–168.

Isaacs, J. D., & Duncan, C. P. Reversal and nonreversal shifts within and between dimensions in concept formation. *Journal of Experimental Psychology*, 1962, **64**, 580–585.

Johnson, P. J. Factors affecting transfer in concept identification problems. *Journal of Experimental Psychology*, 1966, **72**, 655–660.

Kendler, H. H., & Kendler, T. S. Reversal behavior: Some basic issues. *Psychological Bulletin*, 1969, **72**, 229–232.

Polson, P. G., & Dunham, J. L. A comparison of two types of theories of multiple-category concept identification. *Journal of Verbal Learning and Verbal Behavior*, 1971, **10**, 618–624.

(Received July 3, 1972)

Journal of Experimental Psychology
1973, Vol. 98, No. 2, 297–301

DRIVE LEVEL EFFECTS ON THE CONDITIONING OF FRUSTRATION[1]

JEFFREY M. COHEN[2]

Northern Illinois University

Frustration was conditioned in four experimental groups of rats during 12 nonrewarded direct placements which followed 45 rewarded direct placements. Drive level during the rewarded and nonrewarded phases was manipulated by maintaining two groups at a constant drive level (high or low), while the drive level of the two remaining groups was shifted (high to low, or low to high) between the two phases. Then all groups, at their initial drive level, were given hurdle-jumping training to measure conditioned frustration. Both groups which had high drive during the nonrewarded phase jumped faster than comparable low-drive groups, indicating that high drive led to a greater magnitude of conditioned frustration than low drive. Comparisons at each drive level between the constant-drive experimental groups and nonfrustrated control groups indicated that only under high drive was enough frustration conditioned to be detected by the hurdle-jumping measure.

According to the Amsel–Spence theory of frustration (Amsel, 1958; Spence, 1960), unlearned goal responses (R_G) occur in the presence of a goal object (e.g., food), and fractional components of R_G (r_g) become classically conditioned to the situational cues present. When nonreward occurs in the presence of r_g an unlearned response, primary frustration (R_F), is elicited and fractional components of R_F, called conditioned frustration (r_f), become classically conditioned to the cues present.

Of all the studies examining frustration, only a few have provided evidence about the effect of drive (D) level on the magnitude of frustration. Some of these studies have examined primarily the effect of D on r_f; others, on R_F. The studies concerned with r_f have found no effect of D level on the size or occurrence of the partial reinforcement acquisition or extinction effects which were assumed to be determined by frustration (Badia, 1965; Dachowski & Dunlap, 1969; Dunlap, 1969; Dunlap & Dachowski, 1970). On the other hand, the increase in response vigor occurring immediately after the omission of an expected reward has been found to be greater under high- than under low-drive conditions, presumably because D level affects the magnitude of R_F (Dunlap & Frates, 1970; McHose & Ludvigson, 1964). Since it is assumed that R_F is the unconditioned response in the conditioning of r_f, it is difficult to understand how R_F, but not r_f, could be dependent on D level.

It is possible that the negative findings for r_f, with respect to D level, are due to the fact that the partial reinforcement effects did not adequately index the different magnitudes of r_f obtained in the above studies. A different method of measuring r_f has been used by Daly (1969). This procedure involves three phases. Phase 1 consists of a series of trials in which *S*s are placed directly in the apparatus and receive food rewards. The R_G is assumed to be elicited and r_g conditioned on these trials. Phase 2 consists of a series of nonrewarded direct placements with a CS (conditioned stimulus). On these trials, it is assumed that R_F would be elicited and r_f conditioned to the CS and situational cues. In Phase 3 the magnitude of r_f conditioned in Phase 2 is measured by the learning of a new, instrumental response, jumping a hurdle to escape from the frustration-eliciting cues. This procedure, employed in the present experiment, allows comparisons to be made between *S*s with equal D levels during the conditioning of r_g, but with different D levels during the conditioning of frustration. It was expected that the results would show the magnitude of r_f to be dependent on the D level under which it was conditioned, with high D during Phase 2 resulting in a greater magnitude of r_f and, hence, faster hurdle jumping than low D.

---

[1] This article is based on a thesis submitted in partial fulfillment of the requirements for the master's degree at Northern Illinois University. The study was supported in part by Research Grant GB-28197X from the National Science Foundation to W. R. McAllister and D. E. McAllister, whom the author wishes to thank for their assistance throughout the experiment. A report of this research was presented at the 1972 meeting of the Midwestern Psychological Association.

[2] Requests for reprints should be sent to Jeffrey M. Cohen, Psychology Department, Northern Illinois University, DeKalb, Illinois 60115.

METHOD

Subjects and design. Naive male hooded rats ($N = 90$), ranging in age 128–144 days, from the laboratory colony, were used as *S*s. They were divided randomly into six groups, four experimental and two control. The D level, determined by the number of hours since feeding (high D, 22 hr.; low D, 3 hr.), distinguished the experimental groups. These groups were designated HHH, HLH, LLL, and LHL, the three letters referring, respectively, to the D level, high (H) or low (L), during Phase 1 (r_g conditioning), Phase 2 (r_f conditioning), and Phase 3 (hurdle jumping). The two control groups, HHH and LLL, were treated in the same manner as experimental groups HHH and LLL, respectively, except that the control groups did not receive a food reward during Phase 1. Therefore, since r_g would not be conditioned in the control groups in Phase 1, it was assumed that during Phase 2 R_F would not be elicited and r_f not conditioned. In order to differentiate the experimental groups from the control groups, a fourth letter (E or C), referring to whether the group was experimental or control, will be added to each group designation indicated above.

Apparatus. The apparatus consisted of two rectangular boxes placed end to end. The boxes were separated by a common wall containing an opening for a guillotine door. This door, 64 × 84 mm., rested on a hurdle 32 mm. above the floor of the first box and 48 mm. above the floor of the second box. The conditioning of r_g during Phase 1 and of r_f during Phase 2 took place in the first box, which was painted white and was 270 × 114 × 105 mm. Mounted above the conditioning box and serving as its lid was another, hinged box, the bottom of which was covered by sanded Plexiglas. Two bulbs were contained in this box, one to provide a constant illumination of 95 lx., and another to serve as the source of the CS, an increase in illumination from 95 to 1,216 lx. On the r_g conditioning trials, the experimental *S*s received the appropriate number of food pellets in a metal spoon permanently positioned on the wall next to the guillotine door, 54 mm. above the floor. The second box served as the safe box during hurdle-jumping training. It was painted gray and was 248 × 114 × 121 mm. The safe box had a light-source box above it similar to that of the conditioning box except that wire mesh was placed beneath the Plexiglas and the bulb provided a constant illumination of 77 lx. Both boxes had wooden floors, and the floor of the safe box was hinged so as to serve as a floor switch. During Phase 2 the CS was turned on and off manually with a switch. During Phase 3 raising the guillotine door turned on the CS in the conditioning box and started a Hunter Klockounter. Depression of the floor in the safe box terminated the CS and the Klockounter. The Klockounter measured, in units of .01 sec., the latency of the hurdle-jumping response from the conditioning box to the safe box.

Procedure. Ten days prior to the first training day, all *S*s were placed on a feeding schedule which was maintained throughout the initial deprivation period. During this period, at 24-hr. intervals, *S* was removed from its home cage and placed for 1 hr. in another cage with access to food. Water was always available in the home and feeding cages. After 9 days on this regimen, the mean percentage of ad-lib body weight of the groups ranged 83%–84%. On the tenth day following the beginning of deprivation, the *S*s in the low-D condition were fed 4 hr. earlier than usual. One hour before the usual feeding time, *S*s were brought to the experimental room in squads of six and placed in individual holding boxes. Water was available in these boxes throughout the experiment. Each *S* received 7 min. of handling, after which *S* was allowed to explore the safe box of the apparatus for 7 min. The *S* was then placed in another, neutral, box (unpainted) for 7 min., during which time the experimental *S*s were fed 15, and the control *S*s 0, 37-mg. Noyes pellets.

On the following day, Phase 1 began. For the low-D *S*s, the treatments were administered 3 hr., for the high-D *S*s, 22 hr., after eating. All *S*s received 3 trials on the first day, and 6 trials per day for the next 7 days, for a total of 45 trials. A trial consisted of placing *S* in the conditioning box directly over the food cup, which contained either 15 (experimental groups) or 0 (control groups) pellets. The experimental *S*s were removed from the box 10 sec. after consuming the food pellets; the control *S*s, 10 sec. after placement. Throughout the experiment, the apparatus was sponged out after each trial to minimize the effect of odor cues. After the last trial of Phase 1, those *S*s which were to have their D level shifted were fed at the appropriate time so thay would be at their new D level the next day.

Phase 2 consisted of six trials on each of 2 days. These trials were the same as those in Phase 1 except that all *S*s received 0 pellets, and the CS was turned on as *S* was placed in the conditioning box. All *S*s were removed from the conditioning box 10 sec. after placement, at which time the CS was terminated. After the last trial of Phase 2, the feeding time of those *S*s which had had their D level shifted was changed so they would be at their original D level for Phase 3.

Phase 3 consisted of 20 hurdle-jumping trials administered in a single session on the day following completion of Phase 2. The *S* was placed in the conditioning box and 10 sec. later the guillotine door was raised simultaneously with the presentation of the CS. After making the hurdle-jumping response, *S* remained in the safe box for 10 sec. and then was removed to the holding box. If *S* did not jump in 60 sec., it was removed from the conditioning box and a latency of 60 sec. was recorded.

In all phases of the experiment, trials were administered to the *Ss* in squads of six (one *S* from each group) with a variable intertrial interval (range 5–9 min.), determined by the length of time necessary for the other *Ss* to complete a trial.

Results

Hurdle-jumping latencies were converted to speeds by a reciprocal transformation, and means were calculated for blocks of four trials.

In order to determine the effect of D level during Phase 2, a repeated measures analysis of variance was performed on the hurdle-jumping speeds of Groups HHH-E, HLH-E, LHL-E, and LLL-E. The D level during Phases 1 and 3 and the D level during Phase 2 served as the between-*Ss* factors and trial blocks as the within-*S* factor. The performance curves for these groups are presented in Figure 1. Significant effects were obtained for D level during Phases 1 and 3, F (1, 56) = 4.28, $p < .05$; D level during Phase 2, F (1, 56) = 8.93, $p < .005$; and trial blocks, F (4, 224) = 8.00, $p < .001$. The interaction between D level during Phase 2 and D level during Phases 1 and 3 was not significant, F (1, 56) = 1.43, $p > .20$, nor were any of the interactions with trial blocks significant (all ps $> .10$). An inspection of Figure 1, together with the results from this analysis, shows that both groups with high drive during Phase 2 jumped faster than their comparable low-drive groups.

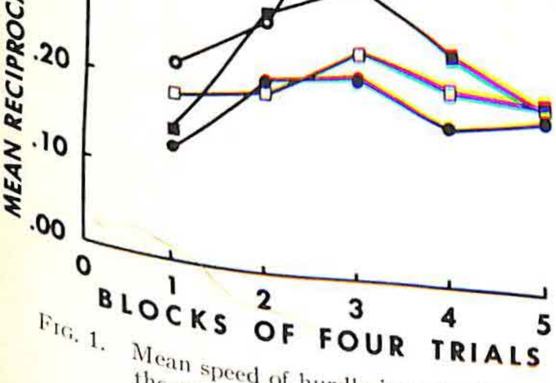


FIG. 1. Mean speed of hurdle jumping for the experimental groups.

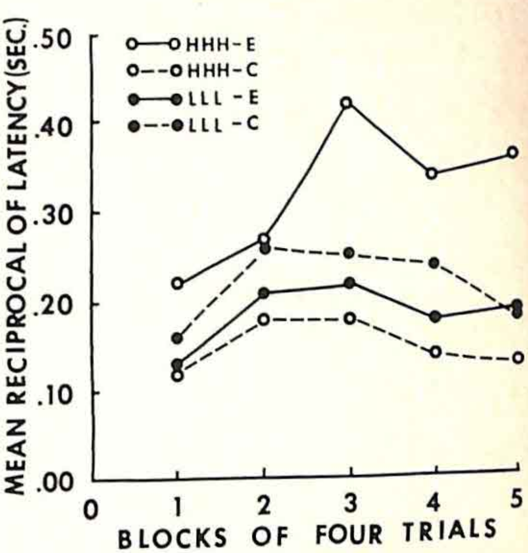


FIG. 2. Mean speed of hurdle jumping for the constant-drive experimental and control groups.

In order to evaluate the amount of frustration conditioned in Groups HHH-E and LLL-E, these groups were compared to their control groups, HHH-C and LLL-C. A repeated measures analysis of variance was performed on the hurdle-jumping speeds with D level (high or low) during the three phases of the experiment and conditions (experimental or control) as the between-*Ss* factors, and trial blocks as the within-*S* factor. The performance curves for the four groups are presented in Figure 2. Only the main effects of conditions, F (1, 56) = 7.55, $p < .01$; trial blocks, F (4, 224) = 7.92, $p < .001$; and the D Level × Conditions interaction, F (1, 56) = 16.73, $p < .001$, were found to be significant. Because of the significant interaction, the effect of conditions at each D level was examined. The results indicated that the experimental and control groups were significantly different under high D, $F(1, 28) = 22.35, p < .001$, but not under low D, $F < 1$.

DISCUSSION

When hurdle jumping took place under high D, the group which also had high D during frustration conditioning (Group HHH-E) performed significantly better than its low-D counterpart (Group HLH-E). This finding ostensibly demonstrates that high D during the conditioning of frustration resulted in a greater magnitude of conditioned frustration than low D. It could be argued, however, that this result was simply due to a stimulus generalization decrement. Specifically, a change in D level between the conditioning of a response and its subsequent elicitation may result in a relative decrease in response strength, because the change in D level changes a portion of the stimulus complex (the drive stimuli) to which the response has been conditioned. Therefore, the inferior performance of Group HLH-E could be attributed to its two shifts of drive level. The first shift was from high D during the conditioning of r_g to low D during its elicitation in Phase 2. This could result in a decreased response strength for r_g during the conditioning of frustration which would result in a decreased magnitude of R_F and r_f. The second shift was from low D during the conditioning of frustration to high D during hurdle jumping. This change in D level might have led to a change in the stimulus complex during hurdle jumping sufficient to result in a decreased response strength for r_f. Therefore, even if low D during the conditioning of frustration did not result in less r_f than high D, it might still result in slower hurdle-jumping speeds.

An evaluation of the validity of this alternative explanation may be made by examining the performance of the groups with low D during hurdle jumping. From a stimulus generalization notion alone, it would be predicted that Group LHL-E, which also had its D level shifted twice, should perform more poorly than Group LLL-E. The results were in the opposite direction and, therefore, suggest that in the present situation a stimulus generalization decrement can be dismissed as the critical factor in interpreting the results. Therefore, it is most parsimonious to account for the differences obtained during hurdle jumping in terms of the effect of D level on the conditioning of frustration. In addition, even though it would be expected that R_F would also be elicited during hurdle jumping and would affect performance, its magnitude would presumably be constant for the groups at a given D level.

The greater magnitude of r_f developed when nonreward was administered under high, as compared to low, D and which led to superior hurdle-jumping performance can be explained in several ways: (*a*) High drive could increase the response strength of r_g and result, upon its disconfirmation, in an increased magnitude of r_f; (*b*) high drive could directly increase the magnitude of R_F and, hence, the magnitude of r_f; or (*c*) both effects could have occurred. Since the magnitude of R_G, at a given D level, was the same during r_g conditioning, the results cannot be due to the effect of D level on the magnitude of R_G. The effects of D level on the other determinants of r_f (r_g and R_F) cannot be differentiated in the present study because they occurred simultaneously during the conditioning of frustration.

A comparison of hurdle-jumping performance at each D level between the experimental groups which did not experience a shift in D level and their respective control groups provides evidence as to whether the experimental procedures were adequate for the conditioning of frustration. Group HHH-E performed significantly better than its control group, HHH-C. This finding indicates that, under high D, nonreward following a series of rewarded trials resulted in frustration which, under the present experimental conditions, was conditioned to the CS and the situational cues. Group LLL-E, however, did not perform in a significantly different manner than its control group, LLL-C. This failure to find a significant difference suggests that under the low-D condition insufficient frustration was conditioned to be detected by the hurdle-jumping measure. There are two possible reasons for this finding: Either the low D during Phase 1 was not sufficient to allow the conditioning of r_g, or the low D during Phase 2 resulted in insufficient r_g or R_F to allow the conditioning of r_f. However, Group LHL-E, which also had low D during Phase 1, demonstrated, in comparison with Group LLL-E, that r_f was conditioned during Phase 2. Therefore, r_g must have been conditioned during Phase 1 under low D.

The results of the present study are contrary to previous studies which have used other measures of r_f in evaluating the effect of drive, but are fully consistent with the results of the studies which have examined R_F. Drive level was shown, in the present study, to have the same effect on r_f as it has been shown to have on R_F; it thus appears that higher D results in an increased magnitude of both R_F and r_f.

REFERENCES

AMSEL, A. The role of frustrative nonreward in noncontinuous reward situations. *Psychological Bulletin*, 1958, **55**, 102–119.

BADIA, P. Effects of drive, reinforcement schedule, and change of schedule on performance. *Journal of Experimental Psychology*, 1965, **69**, 292–297.

DACHOWSKI, L., & DUNLAP, W. P. Frustrative nonreward effects in acquisition under zero hours deprivation. *Psychonomic Science*, 1969, **14**, 113–114.

DALY, H. B. Is instrumental responding necessary for nonreward following reward to be frustrating? *Journal of Experimental Psychology*, 1969, **80**, 186–187.

DUNLAP, W. P. Partial reinforcement phenomena with highly palatable rewards in the absence of food deprivation. Unpublished doctoral dissertation, Tulane University, 1969.

DUNLAP, W. P., & DACHOWSKI, L. Effects of drive and partial reinforcement on extinction both between and within daily sessions. *Psychological Reports*, 1970, **27**, 875–885.

DUNLAP, W. P., & FRATES, S. B. Influence of deprivation on the frustration effect. *Psychonomic Science*, 1970, **21**, 1–2.

MCHOSE, J. H., & LUDVIGSON, H. W. Frustration effect as a function of drive. *Psychological Reports*, 1964, **14**, 371–374.

SPENCE, K. W. *Behavior theory and learning.* Englewood Cliffs, N. J.: Prentice-Hall, 1960.

(Received July 3, 1972)

Journal of Experimental Psychology
1973, Vol. 98, No. 2, 302–307

LEARNING OUTCOMES AND THE LEARNER'S CONSISTENCY SEEKING IN ROTE AND CONCEPTUAL LEARNING[1]

NICHOLAS M. SANDERS[2] AND OVID J. L. TZENG

Pennsylvania State University

An experiment comparing performances in a rote, one stimulus per response task, and two multiconcept tasks, one with the same relevant dimensions across concepts (systematic concepts) and one with no consistently relevant dimensions (unsystematic concepts), revealed that the unsystematic concepts were learned more slowly than the systematic concepts or the one-to-one pairs, which did not differ. However, posttask discrimination of old from similar new stimuli indicated unsystematic-concept task and rote-task *S*s were comparable and more accurate than systematic-concept-task *S*s. Also, a specially devised individual difference measure indicated that learners' consistency-seeking tendencies were facilitative in the systematic-concept condition, debilitative in the unsystematic-concept condition, and not important in the rote task.

Metzger (1958) posed a comparison between performances in rote and concept tasks. He defined *rote learning* as "that type of learning in which the S is required to associate a single response with a single stimulus [p. 226]" and *concept formation* as "that type of learning where S is required to associate the same response to different stimuli [p. 226]." He provided for two types of concept tasks, one in which the same dimensions were relevant for all concepts, *systematic concepts*, and the other in which the particular combination of dimensions relevant for one concept was not relevant for any of the other concepts, *unsystematic concepts*. Using the same stimuli in the rote and multiconcept tasks, Metzger found that (*a*) learning was more rapid in the rote task than in the unsystematic-concept task, but (*b*) learning in the rote task did not differ significantly from that in the systematic-concept task. Metzger suggested that his results could be explained by postulating that one factor, response interference, is always present when a response must be associated with two or more discriminably different stimuli, but that response interference is somewhat offset by a second factor, his "systematic concept formation," when the same dimensions are relevant across all the concepts.

However, Smith, Jones, and Thomas (1963) found that performance on a systematic-concept task was superior to performance on the rote task when *S*s paced the stimulus presentation. Metzger (1958) had used a rapid, 2:2 rate of presentation, which may have inhibited the full operation of the systematic-concept-formation process, as Jung (1968, p. 182) has suggested. However, the Smith et al. systematic-concept task was quite different from that of Metzger, and primary stimulus generalization may have been the process underlying the superior performance of their *S*s in that condition. One purpose of the present study is to compare performances in rote and systematic-concept tasks similar to Metzger's (and others often used in concept-learning studies) when the experiment is S paced.

A second purpose of the present study is to explore one type of learning outcome in the rote, systematic-concept, and unsystematic-concept tasks by determining the accuracy with which *S*s are able to discriminate between stimuli used in the learning task and new similar stimuli after criterion performance in the learning task. If *S*s who learn a systematic-concept task somewhat avoid response interference by associating responses only to the common features of stimuli assigned the same label, generalization of the concepts to the new stimuli would result in misclassifications of the new stimuli. However, it would be expected that the memory of unsystematic-concept-task learners would not differ from rote-task learners, if the unsystematic-concept task leads to a rote, one-to-one learning of stimuli and responses.

---

[1] This research was supported by funds made available from the Advanced Research Projects Agency (ARPA Order No. 1269) through the Office of Naval Research under Contract N00014-67-A-0385-0006.

[2] Requests for reprints should be sent to Nicholas M. Sanders, Department of Educational Psychology, Pennsylvania State University, 201 Social Science Building, University Park, Pennsylvania 16802.

A third aspect of the present study is the determination of effects of differences in learners' task orientations on performances in the three task conditions. While Metzger (1958) seems to have assumed that learners are involved solely in a linking of a particular response to an individual stimulus, many students of concept learning (e.g., Bruner, Goodnow, & Austin, 1956; Laughlin & Jordan, 1967; Peterson & Colavita, 1964) have demonstrated that learners seek, with a variety of strategies, to find consistencies among stimuli grouped together. Although the tasks used by Metzger and in the present study (*a*) were not presented to *S*s with standard concept-learning instructions and (*b*) included only a small number of stimuli with repeated presentations (thus allowing for ease of rote learning in all tasks), it was still possible that some *S*s would actively search for consistencies or rules for the assignment of responses to stimuli in the concept tasks. Other *S*s, of course, might interpret the task, as presented in the instructions, as learning the appropriate response for each stimulus. This latter orientation has not been referred to as a strategy in previous concept-strategy studies. Thus, in the present study, a new measure of learner orientation was devised to differentiate among learners on the consistency-seeking vs. rote-learning dimension. It was hypothesized that consistency seeking should be facilitative in the systematic-concept tasks, but debilitating in the unsystematic-concept tasks in which there is no relevent dimensional consistency among concepts. In the rote tasks, *S*s should be less likely to adopt the consistency orientation because of the independence among stimuli indicated by the different labels for each stimulus.

METHOD

Subjects. Fifty-one college students enrolled in an introductory educational psychology course volunteered for participation in the experiment in order to receive additional credits toward their grade in the course. Data for one *S* in the most difficult condition were excluded due to inability to complete the task.

Design. The *S*s were individually administered a specially devised measure of consistency seeking in learning, followed by two filler tasks. Following these three tasks, the *S* was given one of three experimental tasks: a rote task, a systematic-concept task, or an unsystematic-concept task. After criterion performance on the experimental task, *S* was asked to distinguish between stimuli used in the experimental task and other similar stimuli. All were presented during one session, requiring 60–75 min. for each *S*. The *S*s were assigned to one of the three experimental task conditions in alternating fashion in the order in which they appeared for the session.

Consistency-seeking measure. The materials used for this task included an 11 × 18 in. cardboard, divided by lines into a 4 × 4 set of equal size rectangles. In each rectangle was the letter A, B, D, or E. Each letter was used in four rectangles, though the form of the letter was not the same in any two rectangles; form of the letters differed in case (lower or upper) and in print (script or block). These letters served as labels for 16 $2\frac{1}{2} \times 3$ in. cards, on which there was a square, diamond, triangle, or circle. The geometrical forms varied from card to card in size (with sides of about $1\frac{1}{2}$ in. or $\frac{1}{2}$ in.) and color (black or white) so that no two cards had objects that were exactly the same. The 16 objects were assigned to one of the letter labels on the board by making the four combinations of object size and color dimension correlate perfectly with the letters (e.g., small, white objects were assigned E labels). The labels were arranged on the board so that when the object cards were placed over their appropriate letter labels, the array of cards was not ordered on any object dimension or combination of dimensions. The arrangement of stimuli and labels is shown in Figure 1.

The *S* was shown the board with the object cards placed over their appropriate letter labels so that the labels could not be seen. After a description of the ways in which the objects and labels varied, *S* was told that the task would be to learn which labels went with which objects. He was not told that any characteristics of the objects were related to any characteristics of the labels. Then *E* allowed *S* to uncover up to four labels to study for 15-sec. periods. After 20 such study periods, *S* was asked to recall which labels went with which objects. No feedback for recall accuracy was given.

Scoring for this task was based on the pattern of study choices during the 20 15-sec. periods, not on the basis of recall. It was assumed that if *S* was searching for consistencies, he would select more than one object label to study at a time and the objects

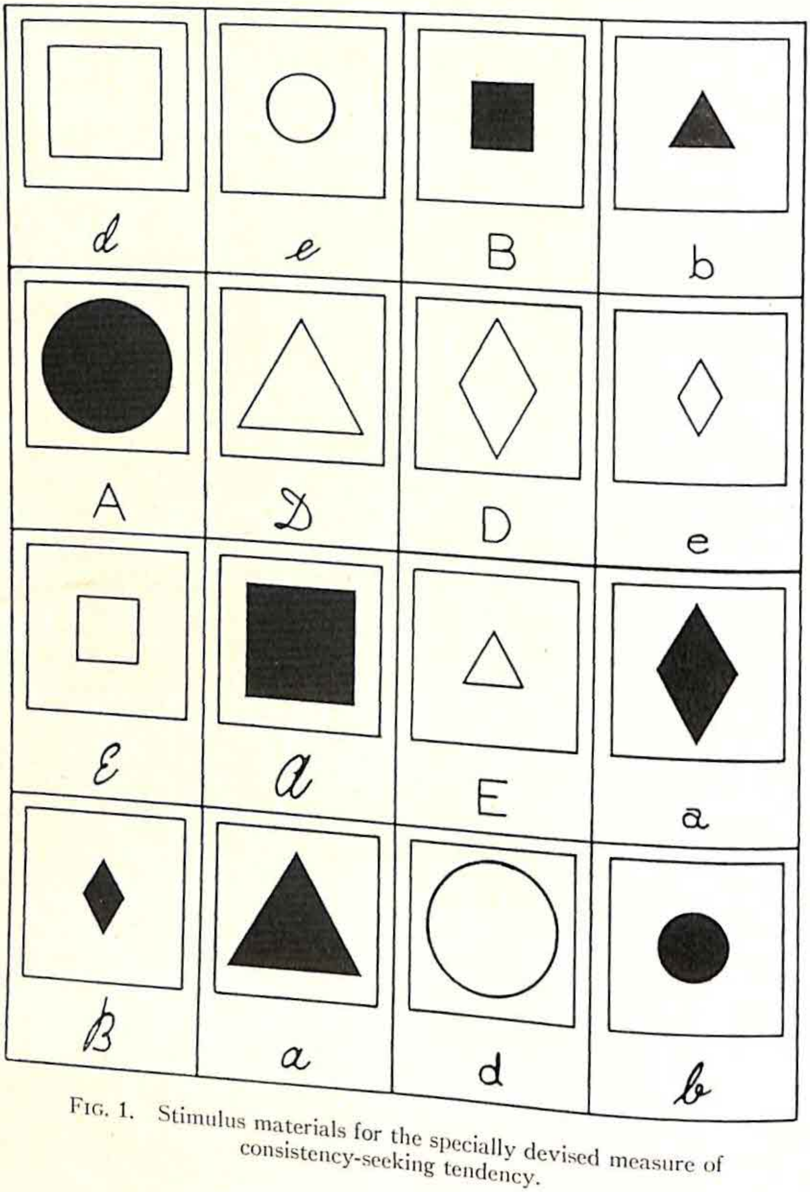


FIG. 1. Stimulus materials for the specially devised measure of consistency-seeking tendency.

selected would be similar to one another on some characteristic or characteristics. Therefore, each study-period selection pattern was recorded and scored by the following procedure: One point was given for study of two objects having at least one feature in common. Two points were assigned for three objects with at least one feature in common. Three points were given for four objects having at least one feature in common. A sum of the points for the 20 study periods was *S*'s score. Thus, a high score on this measure indicates consistency seeking.

Filler tasks. Following the consistency task, *S*s worked on the rote task and a concept task. One half of the *S*s worked on the rote task first, while the other half took the concept task first. These tasks were used to control for and evaluate any set for rote or concept orientations arising from the consistency task.

The rote task involved remembering which of the numbers 1–12 were paired with strings of four *G*s, each string differing from the others by the sequence of capital and small letter *G*s (e.g., g G G g was paired with the number 1). Numbers were assigned randomly to the letter strings, and all pairs were presented simultaneously in numerical order on a piece of typing paper. The *S* was allowed three trial periods of 1 min. of study followed by 1 min. of testing.

The concept task involved the use of 16 stimuli on separate 3 × 5 cards, which were grouped into four sets of four stimuli each. The stimuli were strings of the letters N, O, P, and Q. The letters were always presented in N-O-P-Q order, but each string differed from all others by the particular sequence of uppercase/lowercase forms of the letters. Four concept groupings were identifiable by four colors of the 3 × 5 cards used, and were defined by the case combinations of O and Q (e.g., all sequences with a capital O and a small q were grouped together, and all sequences with both O and Q capitalized were grouped together). Stimuli were described for *S*, and he was instructed that his task was to determine the basis for grouping. He attempted to learn the basis by selecting, one at a time, instances of the groupings for study.

Experimental tasks. The stimuli for these tasks were generated from characteristics of shape (square or diamond), color (red or blue), size (with 3-in. or 1¼-in. sides), and position (right or left). Of the 16 stimuli generated by combinations of these attributes, 8 were selected to provide for comparisons analogous to part of Metzger's (1958) second experiment. The same eight stimuli were used in all three experimental conditions, with differences in response assignments defining the conditions.

In the *systematic*-concept task, the eight stimuli were grouped together on the randomly chosen basis of the four combinations of color and position attributes and were randomly assigned number labels (one, two, three, or four). The *unsystematic*-concept task involved grouping into four groups of two stimuli, so that there were no stimulus dimensions consistently relevant across response categories. In the *rote* task, each stimulus was randomly assigned its own number (1–8). Response labels for stimuli are presented in Table 1.

The *E* read the instructions, which included a description of the dimensions and attributes of the stimuli, the number of stimuli for each number label response, and a description of the *S*-paced stimulus presentation and anticipation-responding procedures. If *S* asked about the basis for assignment of number responses, *E* avoided answering by repeating that the task was to learn which number went with each stimulus. Stimuli were presented one at a time using a Kodak Carousel slide projector. When *S* said a number response, *E* removed the slide and gave the feedback, "Correct, that was a ______" or "No, that was a ______." When *S* was ready for the next stimulus, he indicated so verbally. This procedure continued until *S* had given the correct response for 15 stimuli in succession.

The stimuli were presented in a sequence that prevented stimuli of the same response label from appearing in immediate succession. The eight stimuli were presented in five different subsequences; after the fifth subsequence, stimulus presentation began again with the first subsequence without interruption.

Discrimination task. After criterion performance was attained on the experimental task, *Ss* were told that they would be shown some slides of stimuli presented in the previous task and some new, but similar, stimuli. Their task was to say whether they thought the slide had been presented in the preceding task and, if so, give the appropriate number label for that stimulus. The *Ss* were told that no feedback about correctness would be given during this stage of the experiment.

TABLE 1

RESPONSE LABELS FOR THE EIGHT STIMULI IN THE EXPERIMENTAL TASK

Stimuli[a]	Experimental tasks		
	Systematic	Unsystematic	Rote
SRsr	1	1	3
SRdr	1	3	4
LRdf	2	4	7
SRdl	2	4	1
LBsl	3	3	2
LBdl	3	2	8
SBsr	4	1	5
LBsr	4	2	6

[a] Stimuli are coded as follows: Size—S = small and L = large; color—R = red and B = blue; Shape—s = square and d = diamond; Position—r = right and l = left.

A total of 24 stimuli were presented, the presentation being paced by *S*. Eight of the stimuli were the experimental task stimuli. The 16 additional stimuli involved two presentations of each of the 8 stimuli not previously presented, but which involved variations of the four dimensions used in the experimental stimuli. The order of presentation of the 24 stimuli was random.

RESULTS

Experimental task differences. First, a 3 × 2 factorial analysis of trials to criterion was computed to determine the effects of the three learning task conditions when preceded by either a rote or a concept filler task. The analysis conducted was for unequal cell frequencies and utilized unweighted means. The preceding filler tasks did not produce a significant difference, F (1, 45) < 1, and did not interact with the learning tasks, F (2, 45) < 1. Thus, there is no evidence that a set was developed by the general rote or concept nature of preceding tasks to affect differentially the performance in the learning tasks under study. However, as expected, there was a significant main effect related to the learning tasks, F (2, 45) $= 16.35$, $p < .001$.

Next, Scheffé's multiple comparisons method was used to determine trials to criterion differences between the learning task conditions. As predicted, the un-

TABLE 2

MEANS AND STANDARD DEVIATIONS (IN PARENTHESES) FOR TWO DEPENDENT VARIABLES IN THE THREE LEARNING TASK CONDITIONS

Learning task	n	Criteria	
		Trials to criterion	Discrimination d' score[a]
Systematic concept	17	61.8 (40.28)	1.6 (1.01)
Unsystematic concept	16	128.9 (36.07)	3.0 (1.21)
Rote	17	75.6 (29.61)	3.4 (1.01)

[a] Procedure presented by Elliott (1964).

systematic-concept task was significantly more difficult than either the rote task, $F\ (2, 45) = 18.07$, $p < .001$, or the systematic-concept task, $F\ (2, 45) = 29.92$, $p < .001$. However, the prediction that allowing *S*s to pace themselves would result in more rapid learning in the systematic-concept task than in the rote task was not supported, $F\ (2, 48) = 1.53$, $p > .10$. The relevant means and standard deviations are presented in Table 2.

Discrimination task performances following the original learning were scored using a signal detection procedure described by Elliott (1964), and the resulting d' scores were subjected to a one-way analysis of variance to determine the effects of the three learning task conditions. An unweighted means procedure was used in this analysis, as in the previous ones. The three tasks did result in different discrimination performances, $F\ (2, 47) = 11.49$, $p <$ .001. Scheffé multiple comparisons between means support predictions, with the systematic-concept-task *S*s discriminating less well than either the rote *S*s, $F\ (2, 48) = 12.3134$, $p < .05$, or the unsystematic-concept *S*s, $F\ (2, 45) = 9.1337$, $p < .05$, while the two latter conditions did not significantly differ, $F\ (2, 45) = 2.8110$, $p > .10$. The relevant condition statistics are presented in Table 2.

In order to determine whether the poorer discrimination performance of systematic-concept-task *S*s was the result of generalizing the concepts learned to the new stimuli, each label mistakenly applied by these *S*s to new stimuli was compared with the label that would have been appropriate in generalizing the concepts to the new stimuli. The obtained 71% was compared with a chance level of agreement, 25%, and resulted in a $t\ (16) = 4.67$, $p < .001$.

Individual differences in consistency seeking. The hypothesis concerning relationships of the consistency seeking of learners and learning in the different experimental tasks was tested by comparing the regression of the individual difference measure on trials to criterion for homogeneity of regression among the three learning task conditions. The consistency-seeking measure did differentially relate to learning in the three learning task conditions, $F\ (2, 44) = 6.27$, $p < .01$. As predicted, the greater the consistency-seeking tendency, the more trials were required to learn the unsystematic-concept task ($r = .61$, $p < .01$). Also as predicted, consistency seeking was unrelated to performance in the rote task ($r = -.12$). However, the hypothesized relationship between consistency seeking and trials to criterion in the systematic-concept task did not receive sufficient support; the correlation in this condition was $-.46$, while a correlation of $-.48$ was required for establishing significance at the .05 level.

DISCUSSION

The differences in learning in the three types of learning tasks are in essential agreement with Metzger's (1958) results, though the nonsignificant difference between rote learning and systematic-concept learning was in the opposite direction from that observed by Metzger and was in the same direction as that observed by Smith et al. (1964). Thus, the present study failed to yield evidence that performance in a systematic-concept task can be significantly facilitated, in comparison with rote-task performance, if more time is allowed for the operation of what Metzger has called systematic concept formation.

Taken together, the two previous studies and the present one establish, with different materials and procedures, that learning a set of concepts having different relevant dimensions is much more difficult than learning to classify objects on the basis of common relevant dimensions or learning to label each object individually. However, Metzger's (1958) interpretation of the difficulty of learning unsystematic concepts relied solely on response interference. The present study's positive correlation of individual differences in consistency seeking with trials to criterion in that condition suggests that the tendency on the part of some learners to look for interrelationships among the set of concepts is an additional inhibiting factor.

The present study also included a procedure for revealing particular differences in what is learned in the three types of tasks. The present authors reasoned (*a*) that rote learning would result in accurate discriminations between new and old stimuli, (*b*) that unsystematic-concept-task learners would not differ significantly from rote-task learners, and (*c*) that those *S*s who learned the set of concepts having common relevant dimensions would be less accurate than the others due to generalizing the relevant dimensions to the new stimuli. These predictions were confirmed, indicating an important difference in what is learned among the three tasks.

The present authors also predicted that a tendency on the part of the learner to look for consistencies among stimuli should facilitate performance in the systematic-concept task, inhibit performance in the concept task in which rote processes were appropriate, and be unrelated to rote-task learning. While the latter two predictions were confirmed, consistency seeking was not quite significantly related to performance in the systematic-concept task, though the correlation value was in the appropriate direction. The overall pattern of relationships between the learner's consistency seeking and performance in the three tasks provides some basis for considering individual differences in consistency seeking as an important variable in concept-learning tasks.

REFERENCES

Bruner, J. S., Goodnow, J. J., & Austin, G. A. *A study of thinking.* New York: Wiley, 1956.

Elliott, P. B. Appendix 1—Tables of *d'*. In J. A. Swets (Ed.), *Signal detection and recognition by human observers.* New York: Wiley, 1964.

Jung, J. *Verbal learning.* New York: Holt, Rinehart & Winston, 1968.

Laughlin, P. R., & Jordan, R. M. Selective strategies in conjunctive, disjunctive, and biconditional concept attainment. *Journal of Experimental Psychology*, 1967, **75**, 188–193.

Metzger, R. A comparison between rote learning and concept formation. *Journal of Experimental Psychology*, 1958, **56**, 226–231.

Peterson, M. J., & Colavita, F. A. Strategies, type of solution, and stage of learning. *Journal of Experimental Psychology*, 1964, **68**, 578–587.

Smith, T. A., Jones, L. V., & Thomas, S. Effects upon verbal learning of stimulus similarity, number of stimuli per response, and concept formation. *Journal of Verbal Learning and Verbal Behavior*, 1963, **1**, 470–476.

(Received July 3, 1972)

Journal of Experimental Psychology
1973, Vol. 98, No. 2, 308–315

CONSTRUCT VALIDITY OF RECALL AND RECOGNITION POSTCONDITIONING MEASURES OF AWARENESS[1]

MICHAEL E. DAWSON[2] AND PAUL REARDON[3]

Gateways Hospital, Los Angeles, California

Various types of postconditioning questionnaires have been used previously to measure awareness of the CS–UCS contingency. The construct validity of four of these questionnaires was tested within the context of a masked GSR classical conditioning paradigm. A short recognition questionnaire was found to be a more valid and sensitive measure of awareness than short recall, long recall, or long recognition questionnaires. No evidence of unaware differential conditioning was obtained. However, due to inferential limitations associated with all postconditioning questionnaires, it is recommended that future research develop measures of awareness which can be obtained concurrently with the measures of conditioning.

When repetitive CS–UCS pairings are administered to normal adult humans there may occur an awareness of stimulus contingencies as well as the acquisition of a classically conditioned response. Contemporary conditioning theorists disagree about the role of awareness in human classical conditioning. Some theorists consider awareness to be an undesirable contaminant of "simple conditioning" (e.g., Razran, 1971; Spence, 1966) while others consider it to be an important, perhaps essential, factor in human conditioning (e.g., Tolman, 1959; Woodworth, 1958; and, more recently, Kimble, 1971; Maltzman, 1971). A central issue on which these theorists are divided is whether human classical conditioning can occur when *S*s are unaware of the CS–UCS contingency.

One method used to test for the existence of unaware classical conditioning is to embed the CS–UCS contingency within a masking task. The purpose of the masking task is to prevent, or at least attenuate, the occurrence of awareness. Unfortunately, results of studies which have used this method appear inconsistent. There are reports of both successful and unsuccessful conditioning in the absence of awareness with the masking task technique.

Recent reviews (Dawson, 1973; Grings & Dawson, in press) have indicated that the results of masking task studies are not so inconsistent when one examines the methods used to determine whether awareness was effectively prevented by the masking tasks. Most investigators who found unaware conditioning have employed either a short recall (Lacey & Smith, 1954; Wieland, Stein, & Hamilton, 1963) or a long recall (Fuhrer & Baer, 1969; Wilson, Fuhrer, & Baer, 1971) postconditioning questionnaire. In this context, a short recall questionnaire consists of one or two brief open-ended questions (e.g., "Did you know when you were going to receive the UCS?"), and a long recall questionnaire consists of a series of open-ended questions which become progressively more specific. On the other hand, most investigators who did not find unaware conditioning have used either a short recognition (Dawson, 1970, Experiment II; Dawson & Grings,

---

[1] This research was supported by U. S. Public Health Service Grants MH-18411 and MH-17039, and the Julius R. Wolf Foundation. Computing assistance was obtained from the Health Sciences Computing Facility, University of California, Los Angeles, sponsored by National Institutes of Health Special Research Resources Grant RR-3. The authors thank Fred W. Dunn who assisted in various stages of the research, Barbara Allen who served as one of the interview raters, and Anne Schell and the anonymous consulting editor for many helpful comments. Portions of Experiment I were reported at the meeting of the Society for Psychophysiological Research, St. Louis, October 1971.

[2] Also at California State University at Los Angeles.

Requests for reprints should be sent to Michael E. Dawson, Andrew Norman Research Center, Gateways Hospital, 1891 Effie Street, Los Angeles, California 90026.

[3] Also at University of Southern California.

1968) or a long recognition (Chatterjee & Eriksen, 1960, 1962) postconditioning questionnaire. In this context, a short recognition questionnaire consists of a multiple-choice question from which *S* is to select the correct CS–UCS contingency, and a long recognition questionnaire consists of presenting the CSs and requiring *S* to state whether each was associated with the UCS.

Given that recall questionnaires generally yield results which are different from those of recognition questionnaires, it is reasonable to ask which questionnaire is correct. This raises the issue of construct validity. Construct validity refers to how well a test is able to measure a theoretical construct (in this case, awareness of the CS–UCS contingency). An experimental method of evaluating the construct validity of a measuring device is to employ it in various situations where the construct to be measured will likely vary. If the measuring device is valid, then its scores will vary as the situation varies (Cronbach & Meehl, 1955).

The purpose of the first experiment was to test and compare the construct validity of four postconditioning questionnaire measures of awareness within a masked classical conditioning paradigm. The four questionnaires tested were: short recall, long recall, short recognition, and long recognition. Construct validity was tested by experimentally manipulating the likelihood of awareness among three separate groups and then assessing whether the questionnaires would accurately detect the altered likelihoods. The first group, contingent–instructed (CI), was instructed prior to the conditioning session that there might be a CS–UCS contingency. The second group, contingent–noninstructed (CN), was administered the identical conditioning paradigm but was not instructed about the stimulus contingency. The purpose of these two groups was to determine whether the questionnaires would detect the altered likelihood of awareness due to the instructions. The third group, noncontingent–noninstructed (NN), was not instructed about the stimulus contingency and, in fact, was administered the UCS in an essentially noncontingent manner during the "conditioning" session. The purpose of the third group was to determine whether the questionnaires would detect the absence of awareness due to the noncontingent UCS presentations. In other words, the use of the NN group allows an estimation of the "false positive" rates of each questionnaire.

EXPERIMENT I

Method

Subjects. The *Ss* were 76 paid volunteers solicited through employment offices on college campuses in the Los Angeles area. The data of 16 *Ss* were excluded for the following reasons: 3 due to UCS adaptation, 4 due to procedural errors, and 9 due to inability to correctly report at least $\frac{2}{3}$ of the CSs. The remaining 60 *Ss* were assigned by the method of block randomization to three groups of 20 *Ss* each. Minor deviations from randomness occurred in order to approximate equal sex ratios among the groups, resulting in 11 or 12 females per group.

Apparatus and materials. All *Ss* were tested individually in a sound-attenuated Industrial Acoustics Co. chamber. The same masking task was presented to all *Ss*; it consisted of 30 trials of six tones each. The first tone of each trial was either 950, 1,000, or 1,050 Hz. Each of these three tone frequencies occurred equally often in the first ordinal position. The last five tones of each trial were 800, 950, 1,000, 1,050, and 1,200 Hz. Each of the latter tone frequencies appeared equally often in each of the last five ordinal positions. Several seconds following each trial, *E* said "report" through an intercom system. At that time, *S* reported which of the last five tones was: (*a*) the same pitch as the first tone, (*b*) the highest pitch, and (*c*) the lowest pitch. Since the highest pitch (1,200 Hz.) and the lowest pitch (800 Hz.) tones served as the CSs, an analysis of *S*'s report permitted determination of whether the CSs were correctly reported.

All tones were presented from a tape recorder through a loudspeaker. The tone durations of .5 sec., the intertone intervals ranging 5–13 sec., and the intertrial intervals of 20, 30, and 40 sec. were controlled by Hunter timers during the original taping. The UCS was a pulsating dc shock originating from a Grass S-4 stimulator. The shock was delivered to the volar surface of the first and third fingertips of *S*'s right hand through 1.7-cm.-diam. circular silver electrodes. The UCS duration of .5 sec. and the CS–UCS interval of 1.0 sec. were controlled by Hunter timers.

The galvanic skin response (GSR) was obtained as a dc resistance change recorded with a Grass Model-7 polygraph. Beckman silver–silver chloride electrodes (1 cm. in diameter) were filled with electrode paste and placed on the volar surface of the first and third fingertips of *S*'s left hand. A constant current of 5 μa. was passed through the electrodes so as to not exceed optimal current density.

Procedure. The experimental procedure consisted of six sequential stages:

1. General instructions: While applying the GSR and shock electrodes, *S*s were told that the purpose of the experiment was to determine the effects of "physiological activity" on auditory perception. It was explained that the GSR was the measure of physiological activity and that a brief shock would be administered periodically in order to alter the activity. A description of the "auditory perception test" (masking task) was given.

2. Determination of UCS intensity: Shock intensity was increased in small discrete voltage increments until *S* reported it to be "uncomfortable but not painful." This voltage served as the UCS during the conditioning session.

3. Tone familiarization: The five tones used in the perception task (800, 950, 1,000, 1,050, and 1,200 Hz.) were presented once each in counterbalanced order.

4. Contingency instruction: The contingent–instructed (CI) group was then instructed that "you may or may not find a pattern to the shocks and therefore they may or may not become predictable to you." The other two groups, contingent–noninstructed (CN) and noncontingent–noninstructed (NN), were merely reminded that shock would be presented during the experiment.

5. Conditioning: During conditioning, the 800- and 1,200-Hz. tones served as CS + and CS − (counterbalanced within each group) with 67% reinforcement of CS +. Test trials (nonreinforced presentations of CS +) occurred on Trials 3, 6, 10, 13, 15, 17, 20, 23, 27, and 29. The single largest decrease in skin resistance (transformed to the square root of change in conductance) whose onset occurred 1.0–3.5 sec. following CS onset on test trials was analyzed for discrimination conditioning.

For the NN group, the UCS was paired with each tone equally often during "conditioning." In each block of five shocked trials, the UCS was associated with each tone once. Thus, by the end of the conditioning session, each tone had been paired with the UCS four times. Although there was no actual CS + or CS − for the NN group, the 800- and 1,200-Hz. tones were preselected in counterbalanced order to serve as CS + and CS − for purposes of data analysis.

6. Postconditioning questionnaires: Immediately following the conditioning session, postconditioning questionnaires were administered by an *E* who did not know the group to which *S* belonged. All four questionnaires were presented to each *S*. Half of each experimental group was administered the recall questionnaires first while the other half was administered the recognition questionnaires first.

The long recall questionnaire consisted of 11 questions modeled after those reported by Fuhrer and Baer (1969, pp. 171–172). The questions began at a general level ("Tell me about your experience in the experiment.") and became progressively more specific ("Did you notice any connection between the tones and the shock?"). Following the procedures of Fuhrer and Baer, four judges independently analyzed the content of the transcripts of the long recall questionnaire. Each judge made a rating of either "unqualified aware," "qualified aware," "unqualified unaware," or "qualified unaware." Awareness of the CS–UCS contingency was inferred if there were three ratings of "unqualified aware," or four ratings of "aware" if any judge qualified his rating.

The short recall questionnaire was similar to that reported by Lacey and Smith (1954). These authors used an "extensive interview" but reported that aware *S*s "instantly" verbalized the correct CS–UCS contingency when asked, "Did you know when you were going to be shocked?" As a close approximation to this procedure, Question 7 of the long recall questionnaire was separately analyzed in the present experiment. This question was, "Were you ever able to tell when you were going to be shocked? If so, how?" The criterion of awareness was the same as that for the long recall questionnaire except that judges rated only the answers to Question 7.

The long recognition questionnaire was similar to that used by Chatterjee and Eriksen (1960, 1962). It consisted of presenting each of the five tones used during conditioning and having *S* select one of seven statements about whether each was followed by shock. The order of tone presentation was counterbalanced across *S*s. Awareness was said to exist if *S* selected a statement of higher probability of shock for CS + than for the remaining four tones.

The short recognition questionnaire was patterned after Dawson (1970, Experiment II) and consisted of two multiple-choice questions. The first was "Shock usually followed the (a) highest tone, (b) middle tone, (c) lowest tone, (d) it wasn't systematic, and (e) I couldn't tell." The second multiple-choice question was, "The degree of certainty I have in my above answer is (a) completely uncertain, (b) fairly uncertain, (c) fairly certain, and (d) completely certain." Awareness was inferred if *S* selected CS + on the first question and selected either "fairly certain" or "completely certain" on the second question.

Results

Questionnaire data. Analysis of the postconditioning questionnaire data revealed an interaction between order of presentation and type of questionnaire. This effect was such that the recognition questionnaires detected more awareness when presented early than when presented late, while the recall questionnaires detected more awareness when administered late than when administered early. In other words, recognition was hindered by a prior recall test, while recall was enhanced by a prior recognition test. Postman, Jenkins, and Postman (1948) found the same type of interaction effect between recall and recognition questionnaires while testing for retention of nonsense syllables.

In light of this interaction effect, the present report is limited primarily to the early-presented questionnaire data. This limitation was imposed because only the early-presented questionnaire data are free from the contaminating effects of preceding inquiries. Also, the procedures used to obtain the early-presented questionnaire data most closely parallel those used by previous investigators in that most used only a single questionnaire and thus, in effect, employed an early-presented questionnaire.

Table 1 presents the proportion of *S*s in each of the three experimental groups who were classified as aware with each of the four postconditioning questionnaires. In order to evaluate construct validity, the sensitivity of each questionnaire to the experimental manipulations of awareness was tested. A series of Fisher exact probability tests demonstrated that both the short recall and long recall questionnaires tended to discriminate between the NN and CN groups ($p = .105$) but failed to make any discrimination between the CN and CI groups. The short recognition questionnaire, on the other hand, clearly discriminated between the NN and CN groups ($p = .016$) and also tended to discriminate between the CN and CI groups ($p = .065$). The long recognition questionnaire exhibited poor discrimination between the NN and CN groups ($p = .146$) and between the CN and CI groups ($p = .146$). Thus, of the questionnaires tested, the short recognition questionnaire was the most sensitive to the experimental manipulations of awareness. It was the only questionnaire to significantly discriminate between the NN and CN conditions and the only questionnaire which tended to discriminate between the CN and CI conditions. In signal detection terminology, the short recognition questionnaire exhibited the highest "hit" rate (.90 in the CI group) and the lowest "false alarm" rate (.00 in the NN group).

Also, as can be seen in Table 1, the short recognition questionnaire detected significantly more awareness among *S*s who were administered a CS–UCS contingency (i.e., the CN and CI groups combined) than did either the short recall or long recall questionnaires, χ^2 (1) $= 4.90$, $p < .025$. This greater sensitivity to the presence of awareness is apparently not due to the false positives of the short recognition questionnaire in that it detected no awareness in the NN group. However, the latter conclusion must be qualified by the fact that one does not know what the false positive rate of the short recognition questionnaire would have been in a group treated identically to the CI group in all ways except for the presence of a CS–UCS contingency (i.e., an instructed–noncontingent group).

TABLE 1

Proportion of *S*s Who Exhibited Awareness

Questionnaire	Group		
	Contingent–Instructed	Contingent–Noninstructed	Noncontingent–Noninstructed
Short recall	.30	.30	.00
Long recall	.30	.30	.00
Short recognition	.90	.50	.00
Long recognition	.80	.50	.20

Note. Early-presented questionnaire data only.

Conditioning data. Since previous investigators have suggested that late acquisition trials are the most likely indicators of unaware conditioning (Fuhrer & Baer, 1969; Lacey & Smith, 1954), analyses of the present conditioning data are confined to the last half of the acquisition trials. Also, since the direction of results was clearly specified in advance (i.e., CS + > CS −), one-tailed statistical tests of discrimination conditioning were employed. Unless specified otherwise, classification of aware and unaware groups is based solely upon the early-presented questionnaire data.

Electrodermal discrimination conditioning was demonstrated for the groups judged to be aware with the short recall questionnaire, t (5) $= 2.57$, $p < .05$; the long recall questionnaire, t (5) $= 2.57$, $p < .05$; the short recognition questionnaire, t (13) $=$

2.21, $p < .05$; but not with the long recognition questionnaire. There was no evidence of discrimination conditioning among unaware groups as defined with any of the four postconditioning questionnaires.

In order to test whether the type of questionnaire is a determiner of whether or not one concludes that conditioning occurred without awareness, the GSR magnitude was analyzed among 10 *S*s (4 from the CN group and 6 from the CI group) who were rated unaware by the short recall questionnaire but were judged to be aware with the short recognition questionnaire. The mean GSR magnitude of these *S*s was significantly greater to CS + than CS −, $t\ (9) = 1.91, p < .05$. Thus, there is some evidence that conclusions regarding the occurrence of unaware conditioning may depend in part upon the type of postconditioning questionnaire used to measure awareness.

The conditioning analyses reported thus far have dealt with the issue of whether GSR discrimination conditioning can occur when *S*s are unaware, as measured by various questionnaires, of the CS–UCS contingency. A different but related issue concerns the relationship between the degree of conditioning and the degree of awareness. For present purposes, it was assumed that an *S* who expressed awareness on an early portion of the long recall questionnaire possessed a greater degree of awareness than did an *S* who failed to express it until a late portion of the questionnaire. Therefore, a Spearman rank-order correlation coefficient was computed on the quickness of expressing awareness and the amount of GSR discrimination between CS + and CS −. A significant relationship was found among the 6 *S*s who were aware when the recall questionnaire was presented early, $r_s = .87$, and among the 19 *S*s who were aware on the long recall questionnaire, regardless of its order, $r_s = .51$. Thus, *S*s who express awareness on an early portion of the long recall questionnaire tend to show better GSR discrimination conditioning than do *S*s who express awareness only on a late portion of the questionnaire. Additional correlations performed with the same groups revealed nonsignificant relationships between quickness of expressing awareness and the sum of GSR magnitudes to CS + and CS −. Taken together, these findings indicate that the quickness of expressing awareness is correlated with differential responsivity to the reinforced and nonreinforced CSs but not with general electrodermal responsivity.

Discussion

The major finding of the first experiment was that the short recall, long recall, and long recognition questionnaires were found to be relatively invalid measures of awareness of the CS–UCS contingency. The short recognition questionnaire was the most valid in the sense that it was the most sensitive to the experimental manipulations of awareness. It was the only questionnaire which significantly discriminated between the NN and CN conditions and the only questionnaire which tended to discriminate between the CN and CI conditions. In addition, the short recognition questionnaire detected significantly more awareness among *S*s who were administered CS–UCS pairings than did either of the recall questionnaires.

The poor performance of the long recognition questionnaire, in terms of being sensitive to the experimental manipulations of awareness and in terms of being related to GSR discrimination conditioning, deserves comment. As presently tested, the long recognition questionnaire was found to have two serious shortcomings. First, it exhibited a relatively high false alarm rate (.20 in the NN group). Second, and more seriously, it appears to have exhibited a relatively high false negative rate. There were three *S*s who were clearly aware with the short recall, long recall, and short recognition questionnaires but were unaware with the long recognition questionnaire. Two of these three *S*s rated the tone most similar to CS + (i.e., 950 Hz. instead of 800 Hz.) as being associated with the UCS. Thus, when similar auditory stimuli are employed, the long recognition questionnaire is susceptable to misidentification of CS +. This may not occur when distinctive stimuli, such as individual words, are employed (Chatterjee & Eriksen, 1960, 1962).

The two recall questionnaires, at least when administered early, yielded identical results. However, it should be pointed out that the first experiment relied upon a relatively small sample size (10 *S*s per group who were administered recognition or recall questionnaires early). With a larger sample size, we suspect that the long recall questionnaire might be shown to be more valid than the short recall questionnaire. This hypothesis is based on the observation that when early-presented and late-presented questionnaire data are combined, the long recall measure detected significantly more awareness among *S*s who were administered CS–UCS pairings than did the short recall questionnaire.

Finally, it should be emphasized that the short recall questionnaire consisted of two brief questions selected from an extended interview (i.e., Question 7 of the long recall questionnaire). Although this procedure essentially replicates that of Lacey and Smith (1954), it may be argued that the results of the short recall questionnaire were biased by the preceding interview questions. Therefore, a second experiment was conducted in which the short recall questionnaire was not embedded within a long interview.

EXPERIMENT II

Method

Subjects. The *S*s were 75 paid volunteers solicited through employment offices on college campuses in the Los Angeles area. The data of 15 *S*s were discarded: 7 due to UCS adaption, 2 due to procedural errors, 5 due to inability to report at least $\frac{2}{3}$ of the CSs, and 1 due to excessive basal level.

Procedure. The procedures were the same as those of the first experiment with two exceptions: (*a*) Only the contingent–noninstructed (CN) condition was tested, and (*b*) only short recall and short recognition questionnaires were employed. All *S*s were administered both postconditioning questionnaires. Half of the *S*s were administered the short recall questionnaire first and half were administered the short recognition questionnaire first.

The short recall questionnaire consisted of simply asking, "Did you know when you were going to be shocked?" Any *S* who was not completely negative in answering this question was then asked "to elaborate and be specific about exactly how you knew when you were going to be shocked." Thus, two *S*s who stated "not really" and "not exactly" were asked the second question and then specified the correct CS–UCS contingency. The transcripts of each *S*'s answers to the short recall questionnaire were independently analyzed by two judges. Also, the short recognition questionnaire was modified slightly. Since only 6 of the 60 *S*s in the first experiment selected "middle tone" (and 3 of the *S*s were in the noncontingent group), this choice was eliminated from the short recognition questionnaire of the second experiment.

Results and Discussion

Questionnaire data. The interaction between order of presentation and type of questionnaire found in the first experiment was also observed in the second experiment. That is, the recognition questionnaire detected more awareness when presented first than when presented second, while the recall questionnaire detected more awareness when administered second than when administered first. Therefore, for the reasons presented in Experiment I, the present analysis is limited primarily to the first-presented questionnaire data.

Of the 30 *S*s administered the recall questionnaire first, 10 were rated aware. On the other hand, of the 30 *S*s administered the recognition questionnaire first, 22 were rated aware. This difference was highly significant, χ^2 (1) = 8.10, $p < .01$. These findings confirm those of the first experiment that the short recognition questionnaire detected more awareness among *S*s administered CS–UCS pairings than did a short recall questionnaire.

Conditioning data. Discrimination conditioning of the GSR was demonstrated over the last half of the acquisition test trials for the group judged aware with the short recall questionnaire, t (9) = 2.50, $p < .01$, and the group rated aware with the short recognition questionnaire, t (21) = 2.88, $p < .01$. Discrimination conditioning was not found with either unaware group.

There were nine *S*s who were judged unaware with the short recall questionnaire but who were rated aware with the short recognition questionnaire. The mean GSR magnitude of these *S*s was larger to CS + than CS −, but not significantly, t (8) = 1.04. However, significant discrimination conditioning was demonstrated when the GSR data of these nine *S*s were combined with those of the four *S*s in the CN group of the first experiment who exhibited a similar questionnaire discrepancy, t (12) = 1.84, $p < .05$. Thus, again there is some evidence that whether or not one infers that unaware discrimination conditioning was demonstrated may depend in part upon the type of postconditioning questionnaire used to measure awareness.

GENERAL DISCUSSION

The questionnaire data obtained in the present experiments suggest that a short recognition questionnaire is a more valid and sensitive measure of awareness than are short recall, long recall, and long recognition questionnaires. This conclusion is based on two lines of evidence: (*a*) The short recognition questionnaire was the most sensitive to the experimental manipulations of awareness, and (*b*) the short recognition detected more awareness among *S*s administered CS–UCS pairings than did the other questionnaires. The second line of evidence is based upon the assumption that the false positive rate of the short recognition questionnaire does not exceed that of the other questionnaires. This assumption is partially supported by the lack of false positives exhibited by the short recognition questionnaire in the NN group. However, the false positive rate of an instructed–noncontingent group must be determined in order to fully support this assumption. However, granting this assumption, the data indicate that a short recognition questionnaire is the most valid and sensitive of the questionnaires tested. In addition, it is one of the easiest to administer and can be simply and objectively scored. Therefore, we recommend that future researchers who use postconditioning questionnaires to measure awareness employ a short recognition questionnaire or, even more preferable, that they empirically validate their questionnaire within the context that it is used.

As indicated in the introduction, most previous research which found evidence of unaware discrimination classical conditioning relied upon recall questionnaires. The present data suggest that the results and conclusions of these studies might have been different if the more valid short recognition questionnaire had been employed. For example, in the present research, discrimination conditioning was found among *S*s who were unaware with the short recall questionnaire but were aware with the short recognition questionnaire. If the short recognition questionnaire had not been employed, it might have been erroneously inferred that unaware conditioning was demonstrated among the subsample of *S*s.

The principal conditioning results, at least when related to the more valid questionnaire, were: (*a*) Discrimination conditioning failed to occur in the unaware group, and (*b*) discrimination conditioning did occur in the aware group. The first finding demonstrates that the mere presentation of CS–UCS pairings is not sufficient to establish human GSR discrimination classical conditioning among all *S*s. Some factor or set of factors in addition to physical stimulus contiguity seems necessary for successful conditioning. The second finding is entirely consistent with, but not proof of, the hypothesis that awareness of the CS–UCS contingency is one of the necessary factors for the establishment of human GSR discrimination conditioning.

However, all postconditioning questionnaires have inferential limitations in determining when the awareness occurred, thus making causal inferences about awareness and conditioning inconclusive. That is, postconditioning questionnaires do not demonstrate that awareness was present during the time when differential conditioning was exhibited. It is, for example, possible that the development of conditioning precedes the development of awareness and that a postconditioning questionnaire would not detect this temporal relationship. Only a measure of awareness which occurs concurrently with that of conditioning can detect the temporal relationship between awareness and conditioning. Such a measure of awareness is now being tested in our laboratory.

REFERENCES

CHATTERJEE, B. B., & ERIKSEN, C. W. Conditioning and generalization of GSR as a function of awareness. *Journal of Abnormal and Social Psychology*, 1960, **60**, 396–403.

CHATTERJEE, B. B., & ERIKSEN, C. W. Cognitive factors in heart rate conditioning. *Journal of Experimental Psychology*, 1962, **64**, 272–279.

CRONBACH, L. J., & MEEHL, P. E. Construct validity of psychological tests. *Psychological Bulletin*, 1955, **52**, 281–302.

DAWSON, M. E. Cognition and conditioning: Effects of masking the CS–UCS contingency on human GSR classical conditioning. *Journal of Experimental Psychology*, 1970, **85**, 389–396.

DAWSON, M. E. Can classical conditioning occur without contingency learning?: A review and evaluation of the evidence. *Psychophysiology*, 1973, **10**, 82–86.

DAWSON, M. E., & Grings, W. W. Comparison of classical conditioning and relational learning. *Journal of Experimental Psychology*, 1968, **76**, 227–231.

FUHRER, M. J., & BAER, P. E. Cognitive processes in differential GSR conditioning: Effects of a masking task. *American Journal of Psychology*, 1969, **82**, 168–180.

GRINGS, W. W., & DAWSON, M. E. Complex conditioning. In W. F. Prokasy & D. C. Raskin (Eds.) *Electrodermal activity in psychological research.* New York: Academic Press, in press.

KIMBLE, G. A. Cognitive inhibition in classical conditioning. In H. H. Kendler & J. T. Spence (Eds.), *Essays in neobehaviorism.* New York: Appleton-Century-Crofts, 1971.

LACEY, J. J., & SMITH, R. L. Conditioning and generalization of unconscious anxiety. *Science*, 1954, **120**, 1045–1052.

MALTZMAN, I. The orienting reflex and thinking as determiners of conditioning and generalization to words. In H. H. Kendler & J. T. Spence (Eds.), *Essays in neobehaviorism.* New York: Appleton-Century-Crofts, 1971.

POSTMAN, L., JENKINS, W. O., & POSTMAN, D. L. An experimental comparison of active recall and recognition. *American Journal of Psychology*, 1948, **61**, 511–519.

RAZRAN, G. *Mind in evolution.* New York: Houghton Mifflin, 1971.

SPENCE, K. W. Cognitive and drive factors in the extinction of the conditioned eyeblink in human subjects. *Psychological Review*, 1966, **73**, 445–458.

TOLMAN, E. C. Principles of purposive behavior. In S. Koch (Ed.), *Psychology: A study of a science.* Vol. 2. New York: McGraw-Hill, 1959.

WIELAND, W. F., STEIN, M., & HAMILTON, C. L. Intensity of the unconditioned stimulus as a factor in conditioning out of awareness. *Psychosomatic Medicine*, 1963, **25**, 124–132.

WILSON, R. A., FUHRER, M. J., & BAER, P. E. Differential conditioning of electrodermal responses: Effects of performing a masking task during the interstimulus and intertrial intervals. Paper presented at the meeting of the Society for Psychophysiological Research, St. Louis, October 1971.

WOODWORTH, R. S. *Dynamics of behavior.* New York: Holt, 1958.

(Received July 7, 1972)

Journal of Experimental Psychology
1973, Vol. 98, No. 2, 316–319

EFFECTS OF SIGNAL FREQUENCY ON INCREASE IN REACTION TIME IN A 10-MINUTE AUDITORY MONITORING TASK[1]

HANS-OLOF LISPER[2] AND STIG ERICSSON
University of Uppsala, Uppsala, Sweden

A 10-min. reaction time (RT) task was used to examine if signal frequency affects rate of RT increase. Twelve *S*s participated in four different frequency conditions, corresponding to interstimulus intervals of 4, 8, 12, and 16 sec. It was found that signal frequency affected level of RT performance and rate of RT increase. The RT increase was greatest for the highest signal frequency, a result contrary to findings in experiments on vigilance. As an explanation of this contradiction it was suggested that habituation took place to different aspects of these two types of tasks. Thus, in the RT task, habituation to the signals was the most important factor, while in the vigilance task, habituation to the nonsignal events was more important.

When discussing the effects of signal frequency on vigilance performance, reviewers have reached somewhat different conclusions. Thus, Davies and Tune (1970) proposed an inverted U-shaped relation, i.e., very low and very high frequencies produce poor performance with an optimum somewhere in between, and Stroh (1971) concluded that "performance on a vigilance task is not significantly influenced by signal frequency [p. 9]." The effect of signal frequency on rate of performance decrement is hardly mentioned by these reviewers. However, an examination of reports concerned with signal frequency showed that in about two thirds of these the rate of decrement was the same for all frequencies examined. In the reports where differences in decrement could be observed (Baddeley & Colquhoun, 1969; Bowen, 1964; Goldstein, Johnston, & Howell, 1969; Hawkes & Loeb, 1962; Jenkins, 1958; Johnston, Howell, & Goldstein, 1966; Kappauf & Powe, 1959; Smith, Warm, & Alluisi, 1966), the effects were mostly small and nonsignificant. However, in all the reports mentioned the low frequency or frequencies gave the largest performance decrement. Thus, the effect of signal frequency on rate of vigilance decrement is small. However, the trend over experiments makes it probable that low frequencies favor performance decrement more than high frequencies do.

If signal frequency has the suggested effect on vigilance decrement it is quite contrary to the effects of signal frequency on increase of reaction time (RT) to intense auditory signals. Lisper (1969), having *S*s react to 90-db. auditory signals, found that a high signal frequency (240 signals/hr) gave an RT increase, but that a low signal frequency (12 signals/hr) gave a constant RT over a 3-hr. watch. Since this result, contrary to findings for vigilance, is based on data from only four *S*s it needs further confirmation. Thus, the experiment reported here was done to confirm this surprising effect of stimulus frequency on RT.

METHOD

Twelve university students, six male and six female, aged 19–34 yr., participated as *S*s as part of a requirement in a course in introductory psychology.

The experiment was conducted in a 1.5 × 2 × 2 m. semisoundproof chamber which was dimly lighted during the experiment. The *S*s were comfortably seated in a half-reclining position in an armchair and were instructed to press a microswitch, held in the preferred hand, as rapidly as possible when they heard an auditory signal. This response terminated the signal, which came from a loudspeaker placed in front of and above *S*s. The signals were 1,000-Hz. tones produced by an Oltronix RC oscillator (Type RCO-6K) and having an intensity of 90 db. (re. 0002 dynes/cm^2)

---

[1] This investigation was supported by grants from the Swedish Council for Social Science Research.

[2] Requests for reprints should be sent to Hans-Olof Lisper, Department of Psychology, University of Uppsala, Svartbäcksgatan 10, S-753 20 Uppsala, Sweden.

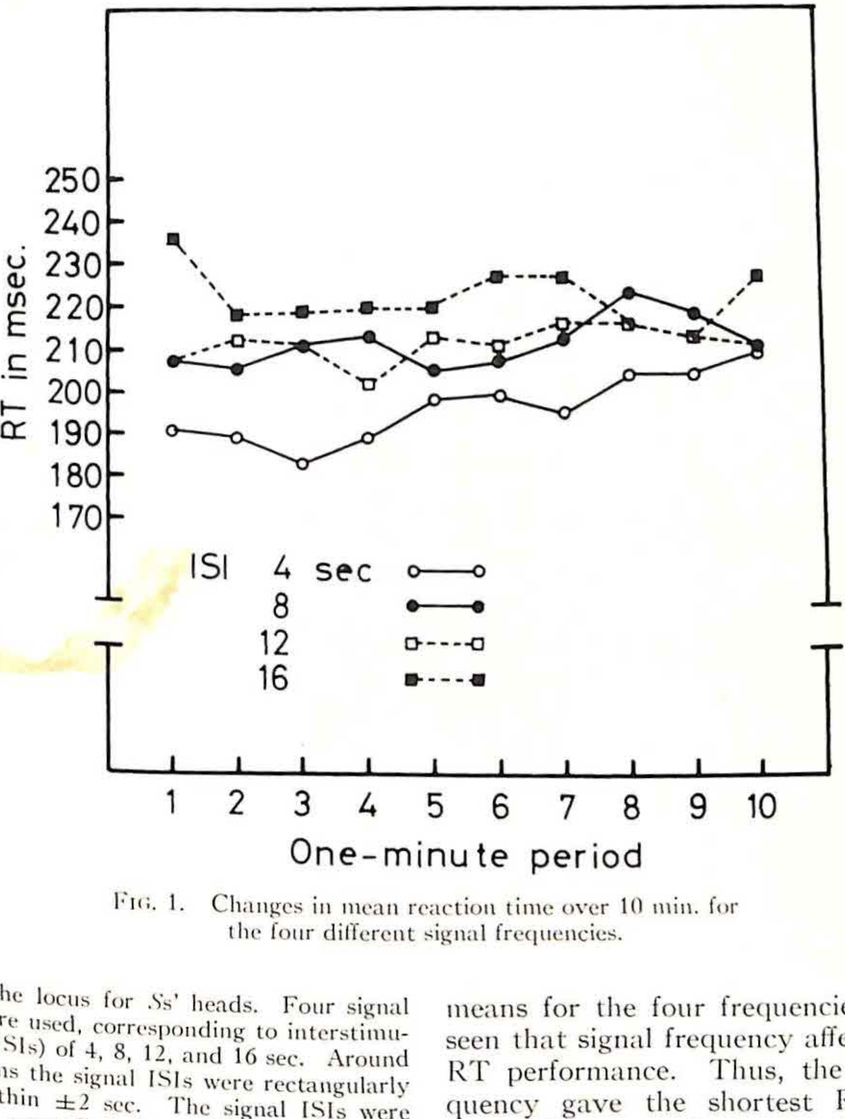


FIG. 1. Changes in mean reaction time over 10 min. for the four different signal frequencies.

measured at the locus for *Ss'* heads. Four signal frequencies were used, corresponding to interstimulus intervals (ISIs) of 4, 8, 12, and 16 sec. Around these ISI means the signal ISIs were rectangularly distributed within ±2 sec. The signal ISIs were automatically controlled by a Tandberg 1200X tape recorder. The *Ss'* RTs were recorded in 1-msec. units by an electronic decade timer, manufactured at the Department of Psychology, University of Uppsala, and were then printed on an ADDO-X printer.

Every *S* participated in all four frequency conditions. The duration of each watch was 10 min., and the watches were separated by a pause of 5 min. during which *S* was requested to leave the chamber. The order of conditions was varied among *Ss* according to three different 4 × 4 Latin squares. Before entering the chamber *Ss* had a 10-min. training period with all four frequencies being represented.

RESULTS

The 10-min. task was divided into 10 blocks of 1 min. each for which the arithmetic means of each *S* and condition were calculated. In Figure 1, showing th groupe means for the four frequencies, it can be seen that signal frequency affected level of RT performance. Thus, the highest frequency gave the shortest RT and the lowest frequency the longest RT, with the RT to the two intermediate frequencies in between. Figure 1 further shows that RT increased most for the highest frequency, less for the intermediate frequencies, and not at all for the lowest frequency. The effect of signal frequency on level of RT performance was significant, $F\ (3,\ 33) = 9.77,\ p < .001$. Furthermore, there was a significant linear main effect of time on task, $F\ (1,\ 99) = 9.50,\ p < .01$, and also the interaction between frequency and linear trend of time on tasks was significant, $F\ (3,\ 297) = 4.72,\ p < .01$.

DISCUSSION

The results confirmed the effect of signal frequency on rate of decrement, inferred from the experiment by Lisper (1969). Thus it is rather certain that signal frequency affects rate of RT increase, and in a direction opposite to what might be the effect in vigilance. A critical question in this connection is that the signal frequencies used in the present experiment were very much greater than those typically found in vigilance studies. However, in the experiment by Lisper the signal frequencies used were within the range of frequencies used in vigilance studies (e.g., Broadbent & Gregory, 1965; Jenkins, 1958; Taub & Osborne, 1968; York, 1962).

A comparison with vigilance may also, on other grounds, seem irrelevant to some readers. The present authors, however, are of the opinion that a strict delimitation of vigilance from nonvigilance tasks is rather irrelevant in relation to theoretical questions. Consequently, it seems relevant to combine findings on performance decrement in vigilance and RT tasks. The combination proposed does not pose any serious problems concerning theories, as there are no specific vigilance theories. The so-called vigilance theories are thus general theories from other areas of psychology, which have been adapted to a quite limited and initially purely applied problem.

Consequently it is reasonable to speculate on how these divergent results on the effects of signal frequency may be combined for vigilance and RT tasks. Mackworth (1968) suggested that performance decrement in vigilance tasks is due to habituation. Thompson and Spencer (1966) stated that one of the characteristics of habituation is that, "Other things being equal, the more rapid the frequency of stimulation, the more rapid and/or pronounced is habituation [p. 18]." Consequently, the effects of signal frequency on the RT increase are in agreement with this habituation characteristic.

Furthermore, it is reasonable to believe that as far as habituation is concerned, the signals are of greater importance in the RT task than in the vigilance task. In the latter task the signals are of little importance compared with the repetitive and high frequency nonsignal events. An experiment by Jerison and Picket (1964) has also shown that in vigilance the frequency of nonsignal events affects rate of decrement in line with the habituation characteristic mentioned above. Although this finding has been contradicted by other results (Taub & Osborne, 1968), it seems reasonable to suppose that decrement in vigilance is due largely to habituation to the nonsignal events and varies little with changes in signal frequency.

It may be that signals in a vigilance task function in accordance with two other habituation characteristics hypothesized by Thompson and Spencer (1966), viz., "Presentation of another (usually strong) stimulus results in recovery of the habituated response (dishabituation) [p. 19]." and "Upon repeated application of the dishabituatory stimulus, the amount of dishabituation produced habituates· · ·[p. 19]." Thus in vigilance, an increase in signal frequency might, up to a certain point, diminish rate of performance decrement. However, it is reasonable to assume that this disinhibitory effect of the signals is very small compared to the massive habituation to the nonsignal background events.

REFERENCES

Baddeley, A. D., & Colquhoun, W. P. Signal probability and vigilance: A reappraisal of the "signal-rate" effect. *British Journal of Psychology*, 1969, **60**, 169–178.

Bowen, H. M. Vigilance as a function of signal frequency and flash rate. *Perceptual and Motor Skills*, 1964, **18**, 333–338.

Broadbent, D. E., & Gregory, M. Effects of noise and of signal rate upon vigilance analyzed by means of decision theory. *Human Factors*, 1965, **7**, 155–162.

Davies, D. R., & Tune, G. S. *Human vigilance performance*. London: Staples Press, 1970.

Goldstein, I. L., Johnston, W. A., & Howell, W. C. Complex vigilance: Relevant and irrelevant signals. *Journal of Applied Psychology*, 1969, **53**, 45–48.

Hawkes, G. R., & Loeb, M. Vigilance for cutaneous and auditory stimuli as a function of intersignal interval and signal strength. *Journal of Psychology*, 1962, **53**, 211–218.

Jenkins, H. M. The effect of signal-rate on performance in visual monitoring. *American Journal of Psychology*, 1958, **71**, 647–661.

Jerison, H. J., & Pickett, R. M. "Vigilance": The importance of the elicited observing rate. *Science*, 1964, **143**, 970–971.

Johnston, W. A., Howell, W. C., & Goldstein, I. L. Human vigilance as a function of signal frequency and stimulus density. *Journal of Experimental Psychology*, 1966, **72**, 736–743.

Kappauf, W. E., & Powe, W. E. Performance decrement at an audio-visual checking task. *Journal of Experimental Psychology*, 1959, **57**, 49–56.

Lisper, H.-O. Classical reaction time and signal rate in a vigilance setting. *Psychonomic Science*, 1969, **17**, 217–218.

Mackworth, J. F. Vigilance, arousal, and habituation. *Psychological Review*, 1968, **75**, 308–322.

SMITH, R. P., WARM, J. S., & ALLUISI, E. A. Effects of temporal uncertainty on watchkeeping performance. *Perception & Psychophysics*, 1966, **1**, 293–299.

STROH, C. M. *Vigilance: The problem of sustained attention.* Oxford: Pergamon Press, 1971.

TAUB, H. A., & OSBORNE, F. H. Effects of signal and stimulus rates on vigilance performance. *Journal of Applied Psychology*, 1968, **52**, 133–138.

THOMPSON, R. F., & SPENCER, W. A. Habituation: A model phenomenon for the study of neuronal substrates of behavior. *Psychological Review*, 1966, **73**, 16–43.

YORK, C. Behavioral efficiency in a visual monitoring task as a function of signal rate and observer age. *Perceptual and Motor Skills*, 1962, **15**, 404.

(Received July 7, 1972)

Journal of Experimental Psychology
1973, Vol. 98, No. 2, 320–327

FOCUSING STRATEGY IN CONCEPT ATTAINMENT AS A FUNCTION OF INSTRUCTIONS AND TASK COMPLEXITY[1]

PATRICK R. LAUGHLIN[2]

University of Illinois

Conservative focusing may be demonstrated to be the most efficient theoretical selection strategy in minimizing the expected number of instances to solution in concept attainment. As a test of this theoretical analysis, four degrees of focusing strategy were induced by instructions: (*a*) unconstrained, the standard instructions where *S* selected any sequence of instances he chose; (*b*) initially constrained, where in each instance *S* had to change one attribute from the initial positive focus instance; (*c*) previously constrained, where in each instance *S* had to change one attribute from the immediately previous instance; (*d*) random, where each instance was selected at random by *S*. As predicted, the order of increasing focusing strategy was: random $<$ unconstrained $<$ initially constrained $=$ previously constrained, and, as a consequence of the degree of focusing strategy, the reverse order of instances to solution was predicted and obtained. The number of attributes and values per attribute were orthogonally varied over three levels of each. As predicted, there was a linear increase in instances to solution with increasing attributes and no effect of increasing values per attribute.

A recent theoretical analysis of selection strategies in concept attainment has demonstrated that conservative focusing is generally the most efficient strategy in minimizing the expected number of instances to solution (Laughlin, in press). This analysis consisted of constructing logical tree diagrams as defined by each of the four selection strategies originally proposed by Bruner, Goodnow, and Austin (1956), as well as by some combined strategies. The mean of the resulting distribution of the number of instance selections to solution for each equiprobable concept may be considered the expected value of a given strategy, and the standard deviation the expected risk or variability of the strategy. An example of this analysis is given in Figure 1, which presents the logical tree diagram for a strategy of conservative focusing which changes one new attribute from a given positive focus instance on each instance selection and makes optimal use of the one verbal hypothesis allowed on each selection. In this example the concept universe consists of a sequence of six ordinal positions (attributes), each of which is a plus or minus sign (values). The given positive focus instance (Miller, 1971) consists of six plus signs, and the instructions indicate that the correct concept is a conjunctive combination of two of these six attributes, such as "one plus and four plus." Hence there are $\binom{6}{2} = 15$ equiprobable two-attribute concepts, and since all 15 involve two plus signs, the plus values are omitted in each of the verbal hypotheses in Figure 1. The first instance selection changes the first ordinal position (attribute), with a probability of 10/15 that the instance will be positive ($P = 10/15$) and a probability of 5/15 that it will be negative ($N = 5/15$). In either case one of the remaining tenable hypotheses is stated. If the instance is positive, the Hypothesis 23 (two plus and three plus) is stated, with a probability of 1/10 that the hypothesis will be posi-

---

[1] This research was supported by a grant from the University of Illinois Research Board and a National Science Foundation Undergraduate Research Participation Grant to the University of Illinois (GY-8744), which supported the participation of Lynette Williamson. Appreciation is expressed to David Drachman, Michael Fishbain, Arnold Grossblatt, Christine Linnas, Andrea Neuberger, Christopher Peterson, William Prosise, Nancy Sailis, Linda Ritzman, and Lynette Williamson for assistance in the collection of the data.

[2] Requests for reprints should be sent to Patrick R. Laughlin, Department of Psychology, Psychology Building, University of Illinois, Champaign, Illinois 61820.

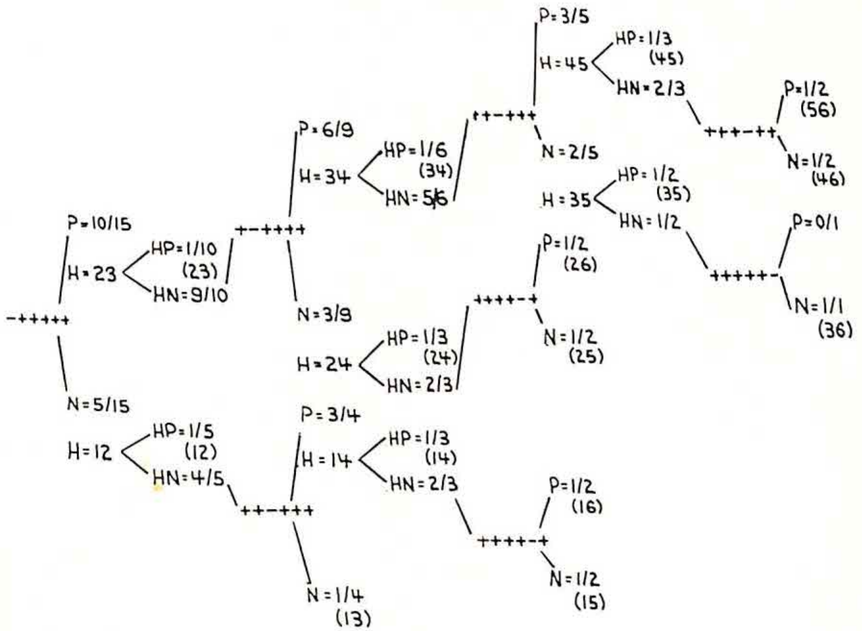


FIG. 1. Logical tree diagram defined by a strategy of conservative focusing with optimal use of verbal hypothesis on each instance selection.

tive (HP = 1/10) and a probability of 9/10 that the hypothesis will be negative (HN = 9/10). Inspection of Figure 1 indicates that there are 12 unique paths of instance selections through the tree, since the three pairs of Concepts 25 and 26, 15 and 16, and 46 and 56 involve single paths of instance selections with the positive or negative status of the last instance determining which member of each pair is the correct concept. The mean of these 12 unique paths which correspond to the 15 equiprobable hypotheses is 2.333, with a standard deviation of .850. Thus, a perfect use of the strategy would lead to an expected value of 2.333 instance selections to solution over a large number of such problems.

Table 1 presents the means and standard deviations corresponding to four strategies for four, six, and eight attributes in the concept universe. These four strategies are: (*a*) successive scanning, in which each hypothesis is tested by a succession of alternating instance selections and verbal hypotheses; (*b*) conservative focusing; (*c*) focus gambling from a succession of two-change selections with correct resolution of

TABLE 1

EXPECTED VALUES OF FOUR SELECTION STRAGETIES FOR CONCEPT UNIVERSES WITH FOUR, SIX, AND EIGHT ATTRIBUTES, TWO RELEVANT ATTRIBUTES, AND A POSITIVE FOCUS

No. of attributes in concept universe[a]	Selection strategy											
	Successive scanning			Conservative focusing			Focus gambling			Generalized tactical		
	N[b]	*M*	*SD*	*N*[b]	*M*	*SD*	*N*[b]	*M*	*SD*	*N*[b]	*M*	*SD*
Four (6)	5	1.800	.748	4	1.500	.500	6	2.000	.816	—[c]	—[c]	—[c]
Six (15)	15	4.267	2.175	12	2.333	.850	15	2.667	.943	10	2.833	1.269
Eight (28)	27	7.259	3.902	24	3.500	1.384	25	3.200	1.058	24	3.583	1.412

[a] Number of equiprobable two-attribute conjunctive concepts appears in parentheses.
[b] Number of unique outcomes, i.e., paired outcomes, counted only once.
[c] With four attributes, selections changing two attributes from the positive focus instance are indistinguishable from either successive scanning or focus gambling.

the ambiguous information from negative instances; and (*d*) a "generalized tactical" strategy, which involves exact selection of any of 30 specified sets of four instances each of which change half of the attributes from the positive focus. This strategy extended the tactical strategy proposed by Johnson (1971) and seems to specify the original Bruner et al. (1956) simultaneous scanning. Finally, the theoretical analysis proposed that these four strategies are points on a continuous dimension of problem solving efficiency, which may be called focusing since all four strategies can be logically reduced to some type of focusing, and conservative focusing is theoretically the most efficient.

The purpose of the following study was to support this theoretical analysis by inducing various degrees of a focusing strategy by four types of instructions. The first instructional condition followed the standard procedure of allowing *S* to choose any sequence of instances he desired, and may be termed *unconstrained* since there were no limitations on the nature or order of allowable selections. In the second, *initially constrained*, condition the instructions required *S* to change one (only) attribute from the known initial positive focus on each selection. It was predicted that this induction of a conservative focusing strategy would result in more use of focusing, and consequently in fewer instances to solution, than the unconstrained condition. Moreover, previous research by Byers, Davidson, and Rohwer (1968) found that *S*s who were instructed to use a conservative focusing strategy performed better than an uninstructed group when a memory aid was not available, and an inverse correlation between the degree of focusing and number of instances to solution under the usual unconstrained selection instructions was reported by both Laughlin and Doherty (1967) and Laughlin and Jordan (1967).

Research on memory effects in conceptual behavior (e.g., Bourne, Ekstrand, & Montgomery, 1969; Bourne, Goldstein, & Link, 1964; Hunt, 1961) has shown that *S*s make progressively less use of previous information with increasing intervening instances. Consequently, the strategy of focusing from the initial positive focus may be decreasingly effective over increasing instances. Thus, in a third, *previously constrained*, condition the instructions required *S* to change one (only) attribute from the previous instance on each selection. It was predicted that performance under this condition (which corresponds to the scoring systems for focusing of Byers, 1967, and Johnson, 1971, which compute focusing relative to the previous rather than the initial instance) would be superior to that under the unconstrained condition, but equal to that under the initially constrained condition, because the scoring rules for focusing credited new information on each instance that changed one attribute not previously proven irrelevant.

The essence of a selection procedure is that *S* chooses whatever sequence of instances he desires. The efficiency of his resulting performance should therefore be compared against a baseline of performance from a random sequence of instances. But, if *E* preprograms this random sequence, the procedure becomes a reception paradigm by definition, and the efficiency of performance will depend upon the information in the resulting sequence. However, the appropriate baseline or control condition may be accomplished by a random selection procedure in which *S* himself selects the instances but by a random procedure. Thus, in a fourth, *random* selection, condition *S* generated each instance by throwing dice to determine the value for each attribute. It was predicted that there would be more use of focusing strategy, and consequently fewer instances to solution, in all three of the other conditions than in this random baseline.

In summary, the predicted order of focusing strategy for the four conditions was: initially constrained = previously constrained > unconstrained > random. As a result of the hypothesized order of focusing, the *reverse* order of number of instances to solution was predicted for the four conditions.

Second, as a test of a possible interaction between the four instruction conditions and task difficulty, the concept universes were orthogonally varied by both four, six, and eight total attributes, and two, three, and four values per attribute. With a positive focus instance, the number of possible hypotheses is determined entirely by the number of attributes, irrespective of the number of values per attribute (Miller, 1971). Consequently, a main effect of number of attributes was predicted, indicating increasing instances to solution with increasing attributes, while no effect of the number of values per attribute was predicted.

METHOD

Design and subjects. The experimental design was a 4 × 3 × 3 × 3 factorial with repeated measures on the last variable. The variables were: (*a*) instructions (unconstrained, initially constrained, previously constrained, or random); (*b*) number of attributes (four, six, or eight); (*c*) number of values per attribute (two, three, or four); and (*d*) successive problems (three of the same type for each *S*). The *Ss* were 144 introductory psychology students of both sexes at the University of Illinois who were fulfilling required experimental participation. All *Ss* were randomly assigned to conditions.

Procedure and instructions. The concept instances were generated by a variation of Huttenlocher's (1962) procedure. The *S* sat at a table before two rows of four, six, or eight 10.16 × 15.24 cm. cards. The top row of cards consisted simply of the numbers 1–4, 1–6, or 1–8 and served to identify the attributes of the cards in the second row by ordinal position. Each attribute (ordinal position) had the value(s) of a plus or minus in the two-value concept universe, and the additional value(s) circle and square was added for the three-value and four-value concept universes. The second row consisted of four, six, or eight plus, minus, zero, or square signs according to the particular concept universe and constituted the known initial positive focus instance of the concept. These signs were all lettered in green ink, so the given positive focus instance was referred to as the "green row." Depending upon the particular number of attributes and values in the condition, there were two, three, or four piles of four, six, or eight plus, minus, zero, and square signs lettered in red ink on 10.16 × 15.24 cards. The *S* generated concept instances by placing whichever one of these red signs he desired under each ordinal position (attribute). When *S* had generated a concept instance or red row, *E* said yes or no to indicate whether it was a positive or negative instance of the concept. The *S* then stated one (only) verbal hypothesis indicating what he considered to be the correct concept, and *E* said yes or no to indicate whether or not the hypothesis was correct. A correct hypothesis solved the problem, while if the hypothesis was not correct, *S* generated a second instance by arranging the appropriate signs in each ordinal position of the red row. This cycle of instance generation, feedback on the instance, one (only) hypothesis, and feedback on the hypothesis continued until *S* verbalized the correct hypothesis, which was the criterion of successful conceptual performance.

The problems were conjunctive concepts with two relevant attributes (e.g., "two minus and three circle"). Four sets of three problems each were randomly selected within each of the nine attribute–value conditions, and the initial instances (green rows) were randomly selected from the subset of possible positive instances for each concept. The instructions thoroughly explained with an example the nature of the task and the conjunctive rule, and explained the procedure indicated above. In the unconstrained conditions no restrictions were placed on the instances that could be generated in the red rows. In the initially constrained conditions the instructions stipulated that each generated instance must vary in only one attribute from the initial positive instance, and in the previously constrained conditions that each generated instance must vary in only one attribute from the previous instance. (There was no explicit requirement that successive instances must necessarily change *new* attributes.) In the random conditions *S* threw a die to determine the value of each attribute on each instance. For example, in the three-value concept universes, a one or two on the die indicated a value of plus on the attribute, a three or four a value of minus, and a five or six a value of zero.

RESULTS

Focusing strategy. Focusing strategy was scored according to two rules. Rule 1: Each instance had to obtain information on a new attribute. Rule 2: The hypothesis had to be tenable. New information was obtained if the instance altered only one attribute not previously proven irrelevant (conservative focusing), or, if the instance altered more than one attribute (focus gambling) it was positive or the ambiguous information on a negative instance was correctly resolved on the next instance by altering only one attribute. With the conjunctive rule and two relevant attributes there were two types of untenable hypotheses: (*a*) Both values of the hypothesis appeared on a previous negative instance, (*b*) either of the other values of either attribute of the hypothesis appeared on a previous positive instance (e.g., the hypothesis "two plus and three square" would be untenable if either a two-minus-or-circle-or-square or three-plus-or-minus-or-circle instance had been positive). The total

TABLE 2

MEAN FOCUSING STRATEGY FOR TOTALS OVER THREE PROBLEMS

Attributes	Values	Condition			
		Unconstrained	Initially constrained	Previously constrained	Random
Four	Two	.46	1.83	1.97	.58
	Three	1.63	2.19	1.67	.19
	Four	1.11	1.91	1.88	.48
Six	Two	1.58	1.99	2.18	.64
	Three	.80	1.75	2.36	.41
	Four	1.39	2.05	1.90	.31
Eight	Two	1.73	2.51	1.62	.88
	Three	1.33	1.92	2.62	.28
	Four	.89	2.33	2.26	.29

Note. Maximum possible focusing strategy is 1.00 on each problem, hence 3.00 for totals over three problems.

number of instances that satisfied both of these two rules was divided by the total number of instances on the problem to obtain a continuous focusing score from 0.00 to 1.00 on each problem. Means for focusing strategy for totals over the three problems are given in Table 2.

The main effect of instructions was highly significant, F (3, 108) = 57.35, $p < .001$, accounting for 33% of the total variance. The order of focusing was random ($M = .45$), unconstrained ($M = 1.18$), initially constrained ($M = 2.05$), and previously constrained ($M = 2.05$). By Newman-Keuls comparisons there was more focusing for each of the unconstrained, initially constrained, and previously constrained, than random conditions (all ps $< .001$); more for both initially constrained and previously constrained than random conditions (both ps $< .001$); and no difference between initially and previously constrained conditions. Thus, the order of focusing for the four conditions was exactly as predicted.

Neither the main effect of attributes nor the main effect of values was significant, F (2, 108) = 1.73; F (2, 108) < 1, respectively. None of the interactions between instructions, attributes, and values were significant.

The main effect of successive problems was significant, F (2, 216) = 11.05, $p < .001$. Trend analysis indicated a significant linear increase in focusing over the three problems, F_{lin} (1, 216) = 21.89, $p < .001$, and Newman-Keuls comparisons indicated more focusing on Problem 2 than on Problem 1 ($p < .01$) with a trend ($p < .10$) to further improvement on Problem 3. The Problems × Instructions interaction was significant, F (6, 216) = 2.10, $p < .05$. The simple main effects of problems for levels of instructions indicated significant improvement over problems for each of the unconstrained ($p < .001$), initially constrained ($p < .001$), and previously constrained ($p < .05$) instructions, while, as would be expected, there was no difference between problems for the random condition.

Number of instances to solution. Means for number of instances to solution for totals over the three problems are given in Table 3. The effect of instructions was significant, F (3, 108) = 10.49, $p < .001$. The order of increasing instances was previously constrained ($M = 11.08$), initially constrained ($M = 11.67$), unconstrained ($M = 16.83$), and random ($M = 18.81$). By Newman-Keuls comparisons there were more instances to solution for random than each of unconstrained ($p < .05$), initially constrained ($p < .001$), and previously constrained conditions ($p < .001$); more for unconstrained than each of initially constrained and previously constrained conditions (both ps $< .001$); and no difference between initially constrained and previously constrained conditions. Thus,

TABLE 3

NUMBER OF INSTANCES TO SOLUTION FOR TOTALS OVER THREE PROBLEMS

Attributes	Values	Condition			
		Unconstrained	Initially constrained	Previously constrained	Random
Four	Two	13.00	10.25	9.50	14.75
	Three	7.25	6.50	7.50	13.50
	Four	13.75	9.75	9.00	9.50
Six	Two	14.50	11.25	10.25	14.25
	Three	19.75	13.25	9.75	23.75
	Four	18.50	14.50	11.25	23.75
Eight	Two	18.75	11.75	18.00	20.25
	Three	23.75	15.50	10.75	25.50
	Four	22.25	12.25	13.75	24.00

the order of instances to solution for the four instructional conditions was exactly as predicted from the degree of focusing predicted and obtained for the four conditions.

The effect of number of attributes was significant, F (2, 108) = 14.67, $p < .001$, indicating more instances to solution for six attributes than four ($p < .001$), and more for eight attributes than six ($p < .01$). The effect of values was not significant, F (2, 108) < 1. None of the interactions between instructions, attributes, and values were significant.

The effect of successive problems was just significant, F (2, 216) = 2.99, $p < .05$. Trend analysis indicated a significant linear decrease in instances to solution over the three problems, F_{lin} (1, 216) = 5.88, $p < .025$, and Newman-Keuls comparisons indicated fewer instances on Problem 3 than 1 ($p < .001$), with no differences between Problems 1 and 2, or 2 and 3. None of the interactions with successive problems were significant.

Correlations between response measures. Correlations between focusing strategy and number of instances to solution over the 36 problems within each of the 12 instruction–attribute conditions (summing over the nonsignificant levels of values) are given in Table 4. As indicated in Table 4, all but one of the 12 correlations was significant at the .05 level or better. Within each of the four instruction conditions, the three correlation coefficients (for four, six, and eight attributes) were converted to z' values, and mean z' values computed for the three attribute conditions. Mean z' values were then compared for the four instruction conditions. There were higher correlations between focusing and instances to solution for both initially constrained, z (105) = 2.86, $p < .01$, and previously constrained conditions, z (105) = 4.02, $p < .01$, than for the random baseline. There was also a higher correlation for previously constrained than unconstrained conditions, z (105) = 2.54, $p < .01$, but no difference between initially constrained and previously constrained conditions, z (105) = 1.17. These analyses thus gave further support to the effectiveness of the initially constrained and previously constrained instructions in inducing focusing strategy and consequently in reducing the number of instances to solution.

TABLE 4

CORRELATIONS BETWEEN FOCUSING STRATEGY AND NUMBER OF INSTANCES TO SOLUTION WITHIN INSTRUCTION–ATTRIBUTE CONDITIONS

Attributes	Condition			
	Unconstrained	Initially constrained	Previously constrained	Random
Four	−.39*	−.70**	−.67**	−.22
Six	−.55**	−.63**	−.78**	−.36*
Eight	−.48**	−.48**	−.64**	−.37*

* $p < .05$.
** $p < .01$.

DISCUSSION

The major finding was the pronounced effect of the four instruction conditions on the use of focusing strategy. The highly significant difference between the random baseline and the unconstrained condition indicates that college students will employ the theoretically maximally efficient strategy of focusing much more than would be expected by chance. The significant differences between both the initially constrained and unconstrained, and previously constrained and unconstrained conditions, indicate that further increases in the use of focusing strategy may be induced both by instructions to vary one attibute from the initial positive focus instance and by instructions to vary one attribute from the previous instance, on each selection. Moreover, these two instructional conditions resulted in identical use of focusing, indicating that the strategy can proceed equally well from the initial instance or the immediately previous instance. This would suggest that scoring systems which determine focusing relative to the initial positive focus instance (Laughlin, 1968), and those which determine focusing relative to the immediately previous instance (Byers, 1967; Johnson, 1971), will give comparable results as long as, in the present study, focusing is credited on each instance that gains information on a new attribute(s) not previously proven irrelevant.

The order of the number of instances to solution corresponded exactly to the predictions from the amount of focusing predicted in each condition. Consequently, it may be argued that the efficiency of concept attainment performance as indicated by the number of instances to solution corresponds directly to the degree to which the strategy of focusing is used in the solution of the problem. This conclusion was further supported by the significant negative correlations between focusing strategy and instances to solution in 11 of the 12 instruction–attribute conditions.

Bruner et al. (1956) originally proposed three other selection strategies in addition to conservative focusing: focus gambling, successive scanning, and simultaneous scanning. If any of these strategies were employed by at least half of the *S*s in the unconstrained condition and were empirically more efficient than conservative focusing as induced in the initially constrained and previously constrained conditions, the number of instances to solution should have been greater in these two conditions than in the unconstrained condition. Thus, the results extend the theoretical analysis of Laughlin (in press) that focusing is the most efficient strategy to an empirical demonstration of the psychological effectiveness of the strategy. Moreover, since conservative focusing makes fewer demands on inference and memory than the other strategies, the results support the interesting conclusion that the theoretically maximally efficient strategy is also psychologically maximally efficient. Finally, since these three other strategies may logically be demonstrated to reduce to varying degrees of focusing strategy, it may be argued that the scoring rules for focusing strategy in this and previous research (e.g., Laughlin, 1968) define a continuous dimension of problem solving efficiency, which may be called focusing, rather than a dichotomous strategy in opposition to scanning.

As predicted, there was a linear increase in the number of instances to solution with increasing attributes in the concept universe. This result corresponds to the demonstration of Miller (1971) that the number of possible hypotheses with a positive focus is a direct function of the number of attributes, and is formalized in Table 1 for the results of logical tree diagrams for each of four strategies for four, six, and eight attributes. However, there was no effect of number of attributes on the use of focusing strategy, indicating that the strategy was used with equal effectiveness for the three attribute conditions. As predicted, there was no effect of the number of values per attribute. Since the problems began with a positive focus instance, the number of possible hypotheses was determined solely by the number of attributes, irrespective of the number of values per attribute (Miller, 1971). Since the typical reception study does not begin with such a positive focus instance, the number of possible hypotheses in the usual reception study is a joint function of the number of attributes and the number of values per attribute, and at least two studies with a reception paradigm have reported a decrease in performance with increasing values per attribute (Battig & Bourne, 1961; Overstreet & Dunham, 1969). Finally, there were no interactions between the number of attributes and the number of values. A previous study of the effects of induced motivation on performance (Laughlin, Chenoweth, Farrell, & McGrath, 1972), also found no interaction between attributes and values in a condition corresponding to the present study, but did find a detrimental effect of the induced motivation of videotaping the experimental session for the most complex attribute–value condition. Thus, the results of the two studies are consistent in the conclusion that attributes and values will interact only under a high level of induced motivation.

Methodologically, the present study used a variation of the Huttenlocher (1962) procedure, in which *S* generated each successive instance by choosing the value of each attribute, rather than the classic Bruner et al. (1956) selection procedure, in which *S* chose each successive instance from an array of all possible simultaneously present instances. While both procedures involve control over the sequence of successive instances by *S* rather than by *E*, they entail obvious differences. For example, in the generation paradigm *S* must actually construct each instance, while in the classic selection paradigm he chooses preexisting instances. Conversely, the classic selection paradigm involves the necessity of locating a particular instance in the total array, which increases in difficulty as a joint function of the number of attributes and values in the concept universe. The generation procedure was used in the present study in order to vary the complexity of the concept universe over a wider variety of attributes and values than would be feasible with the classic selection procedure, where the number of instances is equal to the number of values to the power of the number of attributes. For example, the six-attribute three-value concept universe would have required an array of 729 simultaneously present instances, and the most complex eight-attribute four-value concept universe an array of 65,536. While the generation procedure thus allows the study of a far wider range of complexity in the concept universe, the direct comparison of the two procedures is necessary to determine whether or not they result in comparable principles of human conceptual behavior.

REFERENCES

BATTIG, W. E., & BOURNE, L. E. Concept identification as a function of intra- and inter-dimension variation. *Journal of Experimental Psychology*, 1961, **61**, 329–333.

BOURNE, L. E., JR., EKSTRAND, B. R., & MONTGOMERY, B. Concept learning as a function of the conceptual rule and the availability of positive and negative instances. *Journal of Experimental Psychology*, 1969, **82**, 538–544.

BOURNE, L. E., JR., GOLDSTEIN, S., & LINK, W. E. Concept learning as a function of availability of previously presented information. *Journal of Experimental Psychology*, 1964, **67**, 439–448.

BRUNER, J. S., GOODNOW, J. J. & AUSTIN, G. A. *A study of thinking.* New York: Wiley, 1956.

BYERS, J. L. A note on the calculation of strategies in concept attainment. *American Educational Research Journal*, 1967, **4**, 361–366.

BYERS, J. L., DAVIDSON, R. E., & ROHWER, W. D., JR. The effects of strategy instructions and memory on concept attainment. *Journal of Verbal Learning and Verbal Behavior*, 1968, **7**, 831–837.

HUNT, E. B. Memory effects in concept learning. *Journal of Experimental Psychology*, 1961, **62**, 598–604.

HUTTENLOCHER, J. Effects of manipulation of attributes on efficiency of concept formation. *Psychological Reports*, 1962, **10**, 503–509.

JOHNSON, E. S. Objective identification of strategy on a selection concept learning task. *Journal of Experimental Psychology*, 1971, **90**, 167–196. (Monograph)

LAUGHLIN, P. R. Focusing strategy for eight concept rules. *Journal of Experimental Psychology*, 1968, **77**, 661–669.

LAUGHLIN, P. R. Selection strategies in concept attainment. In R. L. Solso (Ed.), *Contemporary issues in cognitive psychology.* Washington, D.C.: Winston, in press.

LAUGHLIN, P. R., CHENOWETH, R. E., FARRELL, B. B., & MCGRATH, J. E. Concept attainment as a function of motivation and task complexity. *Journal of Experimental Psychology*, 1972, **96**, 54–59.

LAUGHLIN, P. R., & DOHERTY, M. A. Discussion versus memory in cooperative group concept attainment. *Journal of Educational Psychology*, 1967, **58**, 123–128.

LAUGHLIN, P. R., & JORDAN, R. M. Selection strategies in conjunctive, disjunctive, and biconditional concept attainment. *Journal of Experimental Psychology*, 1967, **75**, 188–193.

MILLER, L. A. Hypothesis analysis of conjunctive concept learning situations. *Psychological Review*, 1971, **78**, 262–271.

OVERSTREET, J. D., & DUNHAM, J. L. Effect of number of values and irrelevant dimensions on dimension selection and associative learning in a multiple-concept problem. *Journal of Experimental Psychology*, 1969, **79**, 265–268.

(Received July 10, 1972)

Journal of Experimental Psychology
1973, Vol. 98, No. 2, 328–334

UNIT-SEQUENCE INTERFERENCE IN SHORT-TERM MEMORY: FACILITATION VERSUS INTERFERENCE FACTORS[1]

ANTON K. SABA
University of Maryland

THOMAS W. TURNAGE[2]
Iowa State University

This report provides further evidence for a Meaningfulness × Time interaction in STM whose form is in accord with the unit-sequence interference hypothesis of Underwood and Postman. It also indicates possible reasons for discrepancies among a number of studies which have found, or failed to find, empirical support for the hypothesis of interference. These reasons involve a further consideration of the interplay of interference vs. facilitation effects from intralist vs. extralist sources.

The unit-sequence interference hypothesis predicts that sequences of high-frequency (HF) words should be forgotten faster than equivalent sequences of low-frequency (LF) words, because of opportunities for response competition that increase with word frequency (Underwood & Postman, 1960). Evidence for such effects in short-term memory (STM) comes from a study involving recall of *single* words (Turnage, 1967) and from studies involving recall of *multiple*-word lists (e.g., Postman, Turnage, & Silverstein, 1964; Turnage & McCullough, 1968).

Empirical issues. In four STM studies, Baddeley and Scott (1971) found little evidence for unit-sequence interference in terms of an expected Frequency × Time interaction, in contrast to such evidence reported by Turnage (1967). However, as Baddeley and Scott note, their Experiment I did show some evidence of the predicted interaction, while Experiments II and III have particular method problems. Even so, their Experiment IV is particularly troublesome for the interference hypothesis. Unlike the other three, this study involved a 2–30 sec. retention interval and a large number of *S*s, as did Turnage's study. Still Baddeley and Scott failed to find a reliable Frequency × Time interaction.

Unfortunately, for clarity of interpretation, *intralist* associations and word frequency may have covaried in all four of their studies, especially since their experimental words were randomly selected from a word-count list (see Deese, 1961). Some of these (preexperimental) associations may provide unit-sequence *interference*. Others may provide unit-sequence *facilitation*. The balance of transfer effects would depend on the congruence of the intralist associations with the experimentally, prescribed ones, but the interplay often masks unit-sequence interference (Postman, 1967). However, when intralist associations are minimal, unit-sequence interference should vary directly with word frequency. This is because any facilitation from intralist associations would be reduced, list integrity would be weakened, and the majority of preexperimental associations would then be extralist and interfering (Turnage & McCullough, 1968). Therefore, the first purpose of the present study was to test the following prediction. A Frequency × Time interaction consistent with interference theory would obtain when intralist associations were minimal across the frequency dimension, but results similar to those of Baddeley and Scott (1971) would obtain when number of intralist associations and list frequency were highly correlated.

---

[1] This article is based on a thesis by the first author under the direction of the second author submitted to the faculty of the Graduate School of the University of Maryland in partial fulfillment of the requirements for the M.S. degree. This research was supported, in part, by a grant to the Center for Language and Cognition from the Biomedical Science Support Committee of the University of Maryland. Computation for the project was supported by National Aeronautics and Space Administration Grant NsG-398 to the Computer Science Center at the University of Maryland.

[2] Requests for reprints should be sent to Thomas W. Turnage, Department of Psychology, Iowa State University, Ames, Iowa 50010.

Methodological issues. Turnage (1967) simultaneously presented a to-be-recalled word (HF or LF) and a three-digit number from which *S*s counted to prevent rehearsal. Baddeley and Scott (1971) raise the following objections to this method. First, time sharing between word and number might lead *S*s to take in more information from an LF word vs. an HF word before moving on to the number. Second, this might influence the form of the Frequency × Time interaction found by Turnage. Their first point suggests that LF words might be better learned than HF words. Yet, in terms of estimates from a 2-sec. recall test, the reverse was true in the Turnage study. Saba (1972, Experiment I) investigated their second point by repeating Turnage's single-word study, with the following important modifications in design. The to-be-recalled word was presented alone for 1 sec.; then the three-digit number was presented alone for 1 sec. This successive presentation was to control for the time-sharing artifact. Saba found a Frequency × Time interaction practically identical to that found by Turnage. In both studies, the HF words were recalled better than the LF words at a 2-sec. test, but the reverse was true at a 30-sec. test.

Although potential artifacts deriving from simultaneous presentation did not seem to influence the interaction noted in the Turnage (1967) study, time-sharing problems could have contributed to the Baddeley and Scott (1971) failures to find this Frequency × Time interaction in two of their studies (Experiments I and IV). Here, simultaneous presentation of lists was used, and this procedure could have biased their results against demonstrating unit-sequence interference effects. If *S*s are cautious in emitting an LF word before moving on to the next response, there could be poor processing (learning) of terminal LF items with simultaneous presentation and, hence, apparent better recall of HF lists. Therefore, a second purpose of the present experiment was to investigate interference predictions by using successive, rather than simultaneous, presentation of words in experimental lists.

Theoretical issues. Baddeley and Scott (1971) find Turnage's (1967) single-word study inappropriate because this focuses on item availability while the unit-sequence interference hypothesis focuses on disruption of serial associations *between* items. In addition, they note that multiple-word studies showing an inverse relation between running memory span and word frequency (e.g., Postman et al., 1964) may not reflect unit-sequence interference, but may merely reflect an increased load for order information with respect to HF words since absolutely more of these are recalled. Turnage (1970) has expressed related concerns. Baddeley and Scott used three-item to five-item lists, apparently assuming that short lists reduce such problems of differential availability across the frequency dimension. Therefore, the third purpose of the present study was to investigate unit-sequence interference effects using very short HF vs. LF lists (triads).

METHOD

Materials. The basic experimental materials consisted of a pool of two-syllable nouns varying in frequency of occurrence in the written language (HF vs. LF). For the HF nouns, the number of occurrences in the Thorndike-Lorge L count ranged between 902 and 3,993, while for the LF nouns the number ranged between 1 and 3.

Twenty word triads were constructed from the set of HF nouns so as to be *high* in intratriad associations (HF-H). Each word in an HF-H triad was a high-probability association to at least one other item in that triad. The words involved were drawn from various sources (e.g., Battig & Montague, 1969; Palermo & Jenkins, 1964; Turnage, 1967). Following this, 20 HF triads were constructed so as to be *low* in intratriad associations (HF-L). Finally, 20 LF triads were constructed which were also low in intratriad associations (LF-L). For both HF-L and LF-L triads, each word in the triad rarely called out other words in the triad as associates, according to norms. The HF-L and LF-L triads all came from word pools used by Turnage (1967). Examples of HF-H, HF-L, and LF-L triads were, respectively: MONEY-PAPER-DOLLAR; BUSINESS-COLOR-REASON; and OXIDE-TENURE-STANZA (see Saba, 1972, for additional details).

Design. For each type of triad, the set of 20 was divided into five subsets of four triads each. This division was unsystematic with the restriction that across triads within a subset there were no intertriad associations, there was no duplication of words, and there was low commonality with respect to meaning and sound of items. Each subset of triads was then arranged into a 4 × 4 Latin square. Then, the Latin square for triads was superimposed on

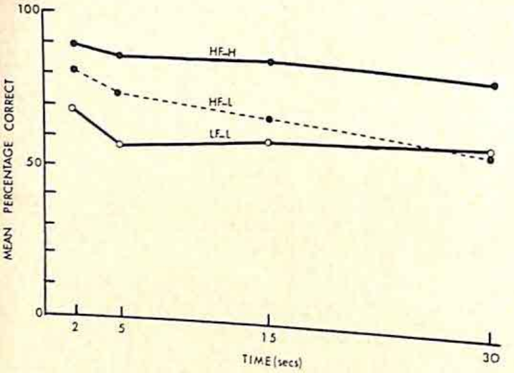


FIG. 1. Mean percentage correct as a function of triad type and time (averaged over four tests). (HF = high-frequency word; LF = low-frequency word; H = high in intratriad associations; and L = low in intratriad associations.)

4 × 4 Latin squares for the retention intervals (2, 5, 15, and 30 sec.). This gave five different Graeco-Latin squares for each type of triad (HF-H, HF-L, and LF-L). Each row of the Graeco-Latin squares designated the type of triad to be recalled and the order of testing with respect to retention intervals for a given group of *Ss*. Thus, each *S* had four different retention tests with four different triads of the same type. A three-digit number was assigned to each triad by a table of random numbers. The *Ss* were to count from this number to control rehearsal during the retention interval.

The counterbalancing procedures used permitted the blocking of *Ss* for data analyses such that within each block there appeared one *S* whose first recall test was at 2 sec., another whose first recall test was at 5 sec., another whose first recall test was at 15 sec., and another whose first recall test was at 30 sec. Correspondingly, within each block there also appeared one *S* whose second test was at 2 sec., another whose second test was at 5 sec., and so on to the fourth recall interval and the last recall test. This blocking procedure was used to analyze the effects of repeated testing (0–3 prior tests) on recall over the 30-sec. retention interval. Thus, triad type was a between-*S* (blocks) variable; while time, serial position, and number of prior tests were within-*S* (blocks) variables.

The *Ss* were assigned in groups of four to experimental conditions at random as they entered the laboratory, with the restriction that every three groups of *Ss* were tested with all three kinds of triads. The design was quite comparable to Turnage's (1967); however, the following modifications should be noted: (*a*) Triads were used instead of single words, (*b*) group testing was substituted for individual testing in accord with the Baddeley-Scott (1971) procedures, (*c*) words in a triad were presented singly, followed by a three-digit number, as constrasted to simultaneous word–number presentation, (*d*) stricter criteria were adopted to control for rehearsal, (*e*) the counting task was written and paced, as contrasted to oral and unpaced, and (*f*) instructions to *Ss* were tape-recorded.[3]

Procedure. Presentation of stimulus material was serial and at a 1-sec. rate by means of a rear-view screen. The sequence of events in a trial was as follows: *E* rang a bell which served as a "ready" signal. This was followed by the first word of the triad, then by the next word of the triad, and then by the last word of the triad. Each word appeared for 1 sec., and *Ss* read silently. Then a three-digit number appeared for 1 sec. and *Ss* wrote it down and began writing their backward count by 3s for the duration of the retention interval, paced by a metronome preset at 72 beats/min (1.2 beats/sec). The screen was blank during this interval. Then, a buzzer signaled *S* to write the words just presented in their correct serial order. Then, 15 sec. later, a bell sounded indicating the end of that trial. At this signal, *Ss* turned to the next page of their booklet. A 10-sec. intertrial interval then followed.

The *Ss* were carefully monitored to see if they were following instructions during their counting task. In addition, two criteria for rehearsal were adopted. The first involved any indication—either through direct observation or through posttest inquiry—that *Ss* deliberately rehearsed. The second involved a check to see if *Ss* had produced a "reasonable" number of counts during the retention interval. A prior study involving 32 *Ss* indicated that the mean number of backward counts by 3s that could be produced from a random three-digit number was about 29 counts/min, with a range of 26–34. On the basis of these data, cutoff points were established as follows. The *Ss* were required to give at least 1 count/2 sec; 2 counts/5 sec; 5 counts/15 sec; and 10 counts/30 sec. Those *Ss* failing to meet the above criteria in any trial were dropped from the study.

Subjects. The total number of *Ss* used was 240. An additional 32 *Ss* took part in the experiment but were replaced for various reasons such as failure to meet rehearsal criteria or failure to follow instructions. There were approximately as many females as males in the experiment.

Since the *Ss* were tested in groups of four, each of these groups was considered as a single *S* in the general analysis. This gave a total of $n = 20$ for each of the three types of triads. When these *ns* were blocked, as described in the Design section, this gave a total of five blocks for each type of triad for the final data analysis. Variation among blocks was then used as an estimate of experimental error.

RESULTS

Serial recall. Figure 1 shows serial recall as a function of triad type and time expressed in terms of mean percentage correct to facilitate comparisons with the data of

---

[3] Thanks are due to Patricia Stine for her services here.

Turnage (1967) and of Baddeley and Scott (1971).

An analysis of variance performed on correct serial recall, items correctly spelled and in the correct serial position, indicated that the Triad Type × Time interaction was reliable, F (6, 36) = 2.66, $p < .05$. Further, this interaction for only HF-L and LF-L triads was also reliable, F (3, 36) = 3.97, $p < .05$. This Frequency × Time interaction for HF-L and LF-L triads is of the same form as that found by Turnage for single HF vs. LF words and is consistent with the hypothesis of unit-sequence interference in STM. That is, although HF-L triads were better learned than LF-L triads, according to the 2-sec. recall test, they were forgotten faster over time. The inverse relationship between original learning and final recall levels permits this conclusion concerning differences in forgetting rates, even though degree of learning was not equated (Underwood, 1964).

Figure 1 also indicates that when HF-H vs. LF-L triads are compared, the general findings of Baddeley and Scott (1971) obtain. That is, there appears to be little difference in rate of forgetting as a function of frequency. This is contrary to the hypothesis of unit-sequence interference in STM. However, this outcome could be due to a confounding of frequency and number of intralist associations (Turnage & McCullough, 1968) as well as to the size and the direction of the uncontrolled differences in degree of learning (see Underwood, 1964). Both these factors may also have been involved in the Baddeley-Scott failures to find support for interference predictions. Correspondingly, in a comparison of HF-H vs. HF-L triads, the unit-sequence interference hypothesis would seem to predict faster forgetting of serial order for the HF-H items because of more potential intralist competition. However, Figure 1 shows a strong, direct relationship between learning and recall levels for these triads. Thus, unambiguous conclusions concerning the influence of intralist associations on rate of forgetting are difficult to draw here (see Underwood, 1964).

As Figure 1 suggests, the main effect of triad type was reliable, F (2, 12) = 33.59, $p < .001$. There were no reliable interactions involving number of prior tests, triad type, and time as common components (all Fs < 1.40). However, important differences in degree of original learning may have influenced these particular results. As expected, time was a reliable source of variance, F (3, 36) = 6.65, $p < .005$. However, the general rate of forgetting indicated for our triads appeared to be slower than that found by Baddeley and Scott (1971, Experiment I) for their triads. This is probably due to the longer study time used here, but it may reflect differences in experimental materials across studies.

Repeated testing. Number of prior tests was a reliable main effect, F (3, 36) = 8.23, $p < .001$. The correct recall means for Tests 1–4, respectively, were 3.27, 2.90, 2.89, and 2.78. This systematic decline in recall level suggests that interference effects from prior testing were not complete by the second test (see Turnage, 1967). However, there was no evidence that the general form of the Triad Type × Time interaction shown in Figure 1 was due, primarily, to repeated testing effects. Table 1, which shows mean percentage correct on the first test, gives more evidence on this point. Note here that the trend for HF-L triads fails to indicate that, when interference from prior testing is precluded, all forgetting curves reach an asymptote in about 5 sec. (see Baddeley & Scott, 1971).

Serial position. As expected, the main effect of serial position was reliable, F (2, 24) = 11.48, $p < .001$. The means for Positions 1–3, respectively, were 3.08, 2.80, and 3.00. Thus, the usual primacy–recency

TABLE 1

MEAN PERCENTAGE CORRECT AS A FUNCTION OF TRIAD TYPE AND TIME ON THE FIRST OF FOUR RETENTION TESTS

Triad type	Time of retention test (in sec.)			
	2	5	15	30
HF-H	93	90	88	88
HF-L	87	83	83	70
LF-L	73	68	78	77

Note. HF = high-frequency word; LF = low-frequency word; H = high in intratriad associations; and L = low in intratriad associations.

curve for serial learning obtained. However, since there were no reliable interactions involving serial position as a component (all Fs < 1.31), there was no direct evidence that time-sharing problems with respect to serial position influenced the form of the Frequency × Time interactions found by Baddeley and Scott (1971).

Word availability. Recall without regard to correct serial position (lenient scoring) was analyzed as a measure of word availability. Triad type was a reliable main effect, $F(2, 12) = 31.63, p < .001$, as might be expected from studies of free recall (see Deese, 1961). Mean number of words recalled (out of a possible 12) for HF-H, HF-L, and LF-L triads were, respectively, 10.85, 9.35, and 8.08. Differential availability of HF vs. LF words and the lack of a reliable Triad Type × Frequency interaction, $F(6, 36) = 1.55$, require further consideration of the hypothesis of disproportionate memory load for order information (Baddeley & Scott, 1971) as a potential explanation for the data in Figure 1. We will return to this point. The only other reliable sources of variance here were time, $F(3, 36) = 5.21, p < .01$, and number of prior tests, $F(3, 36) = 8.06, p < .001$. Means at 2, 5, 15, and 30 sec. were, respectively, 9.98, 9.32, 9.20, and 9.20; while for Tests 1–4, they were 10.25, 9.20, 9.25, and 9.00, respectively.

DISCUSSION

Baddeley and Scott (1971) conclude that unit-sequence interference is not an important determinant of forgetting in STM on (*a*) empirical, (*b*) methodological, and (*c*) theoretical grounds. In view of our present results, their general conclusion appears open to question.

Empirical issues. The Frequency × Time interaction predicted from interference theory occurs when word frequency and number of intralist associations are not highly correlated. (Compare HF-L vs. LF-L triads in Figure 1.) When they are highly correlated, results consistent with the negative findings of Baddeley and Scott (1971) occur. (Compare HF-H vs. LF-L triads in Figure 1.) Note also that the *form* of the interaction for HF-L vs. LF-L triads over our 15-sec. interval is similar to the "nonsignificant" interaction for triads found by Baddeley and Scott (1971, Experiment I) over their 16-sec. interval. Since it is likely that their shortest lists (triads) were also lowest in number of intralist associations (see Deese, 1961), their findings in Experiment I seem consistent with ours for HF-L vs. LF-L triads, projected over the entire 30-sec. retention interval.

Turnage (1967) and Saba (1972, Experiment I) found no forgetting for their single LF words over a 30-sec. interval. Both studies showed a slight reminiscence effect for these words. Figure 1 shows that there was also little forgetting for the LF-L triads formed from these *same* LF words and that a slight reminiscence effect also obtained. These results contrast to the considerable forgetting found by Baddeley and Scott for their LF items. This may reflect the fact that their LF words have an L-count range three times that of the present LF words. Thus, more unit-sequence interference and, hence, more forgetting would be expected for their items. This difference is also important because empirical outcomes in apparent support of the unit-sequence interference hypothesis (e.g., Postman et al., 1964; Turnage, 1967) depend on the particular frequency values selected for simple HF vs. LF comparisons (see Turnage & McCullough, 1968).

Methodological issues. There is no evidence that the form of the Frequency × Time interaction reported by Turnage (1967) was an artifact of simultaneous presentation of words and digits as suggested by Baddeley and Scott (1971). A comparable interaction obtains when successive presentation is used for single-HF vs. single-LF words (Saba, 1972, Experiment I) or for HF-L vs. LF-L triads (see Figure 1). Correspondingly, the present study does not indicate that simultaneous presentation of multiple-word lists affected the form of the (nonsignificant) Frequency × Time interactions reported by Baddeley and Scott. However, there is evidence that confounding of word frequency and intralist associations is an important consideration in interpreting data relevant to the unit-sequence interference hypothesis.[4] When frequency and intralist associations are highly correlated, results comparable to those reported by Baddeley and

---

[4] In this regard, the Baddeley-Scott word lists were recently made available to us (A. D. Baddeley, personal communication, September 21, 1972). Two examples of word lists from their critical Experiment IV are: MIX-BOY-AIR-JOB-FLY (HF items) vs. ROE-TAR-VIE-WED-ZOO (LF items). These lists suggest that their HF items could also be easily coded as sentences.

Scott obtain. When this correlation is reduced, results *consistent* with the unit-sequence interference hypothesis obtain (see Figure 1).

Our lenient scoring (availability) analysis indicated that HF words were better retained than LF words. Except for their Experiment III, the same was probably true in the Baddeley-Scott (1971) studies. In general, they note that such availability differences may result in a disproportionate increase in memory load for order information across the frequency dimension. This, in turn, may result in spurious unit-sequence interference effects (also see Turnage, 1970). Their use of short experimental lists, in contrast to the much longer lists used by Postman et al. (1964), was apparently due to such considerations. However, serial disruption may still vary with word frequency even when item availability is formally equated by presenting the to-be-ordered words during the recall test (Turnage & McGinnies, in press). In addition, availability differences favoring HF items do not necessarily bias results in favor of interference predictions. Correct guessing of serial position for highly available HF words could bias results in the opposite direction, especially with short lists.

In any case, since the five-word lists used by Baddeley and Scott (1971, Experiment IV) did not appear to result in fortuitous unit-sequence interference effects due to greater availability of HF words, it is unlikely that our present three-word lists would do so. Note in this regard that the most available words from HF-H triads were also the best recalled serially. Of course, this could be because HF-H triads were best learned. Hence, high levels of serial recall would be expected. However, HF-L triads were also better learned serially compared to LF-L triads, but they were not better recalled serially presumably because of interference effects.

Theoretical issues. Baddeley and Scott (1971) suggest that Turnage's (1967) single-word study was inappropriate as a test of unit-sequence interference because it focused on item availability and not on serial disruption in recall. In this regard, characterization of Turnage's experiment as a single-word study may be somewhat misleading. Each *S* in his study learned and recalled a total of four test words (HF or LF), one at a time, over a 30-sec. interval. When each test beyond the first is considered, experimental words can be viewed as belonging to a set (or list) of two, three, or four experimental items, presented and recalled in a particular serial (test) order. Conceivably, failures of recall here could involve unlearning and recovery of competing responses with respect to serial (test) position; and this might vary directly with word frequency in accord with the unit-sequence interference hypothesis. Correspondingly, since proactive inhibition (PI) for terminal items in lists appears to increase with list length (See Postman et al., 1964), PI might also cumulate because of increasing set length with each successive word recalled. Further, such PI might vary with word frequency (see Underwood & Ekstrand, 1967). Turnage found a Frequency × Time interaction (averaged over all four tests) and a Frequency × Number of Prior Tests interaction whose forms seemed in accord with interference predictions. Note also that the pool of HF and LF words involved was the same as that used to form the present HF-L and LF-L triads; hence, Turnage's words sets might be characterized as HF-L and LF-L "lists." Thus, the interaction for HF-L vs. LF-L triads shown in Figure 1 seems quite consistent with the Frequency × Time interaction found by Turnage for "single" words.

Extralist vs. intralist interference. There seems to be reliable evidence for memory failures in STM that appear in accord with the unit-sequence interference hypothesis. However, there is little convincing evidence that unlearning and recovery of response competitors produce the forgetting noted, even when strong unit-sequence interference effects are presumed to be operating in STM (Turnage & Steinmetz, 1971). Therefore, an interference explanation which does not require the unlearning and recovery assumption seems necessary. In this regard, consider the learning and recall levels for triads shown in Figure 1. These might be viewed in terms of the A-B, A-C interference paradigm, where A-B refers to preexperimental associations and A-C to experimentally prescribed ones.

In the present study, list integrity for HF-H triads should be high and extralist A-B interference should, therefore, be minimal because of intralist associations of both the A-B and A-C variety (see Deese, 1961). Moreover, some of the intralist A-B associations might correspond to prescribed A-C associations, thereby facilitating serial learning and recall (Postman, 1967). Thus, since general interference for A-C associations from A-B sources (either intralist or extralist) would be greatly attenuated, learning and retention should be good for HF-H triads. In contrast, since both HF-L and LF-L triads have few intralist A-B associations, list integrity here should be poor.

Consequently, interference from extralist A-B associations should be more likely, and since there are more A-B associations for HF words, such interference on A-C associations should vary directly with word frequency. Availability differences favoring HF words might facilitate serial learning of HF-L vs. LF-L lists, but be offset by increasing interference from extralist A-B sources as list differentiation becomes poorer over time (See Figure 1). Parenthetically, this analysis of extralist vs. intralist sources of interference does not seem too radically different from one offered by Underwood (1954) with respect to similarity variables as sources of transfer in learning and retention.

Presumably, extralist interference effects would be enhanced by, but not depend upon, repeated testing (see Table 1). Such an analysis also suggests a prediction not usually made from the unit-sequence interference hypothesis: Repeated testing should produce more interference for LF-L lists than for HF-H lists because of greater list integrity for the HF-H lists. Due to problems involving differences in degree of original learning, the present data are ambiguous on this point. However, Underwood and Ekstrand (1967) reported trends indicating more PI due to repeated testing for LF-L than for HF-H lists in a paired-associate study.

REFERENCES

Baddeley, A. D., & Scott, D. Word frequency and the unit-sequence interference hypothesis in short-term memory. *Journal of Verbal Learning and Verbal Behavior*, 1971, **10**, 35–40.

Battig, W. F., & Montague, W. E. Category norms for verbal items in 56 categories: A replication and extension of the Connecticut category norms. *Journal of Experimental Psychology Monograph*, 1969, **80** (3, Pt. 2).

Deese, J. From the isolated verbal unit to connected discourse. In C. N. Cofer (Ed.), *Verbal learning and verbal behavior*. New York: McGraw-Hill, 1961.

Palermo, D. S., & Jenkins, J. J. *Word association norms*. Minneapolis: University of Minnesota Press, 1964.

Postman, L. The effect of interitem associative strength on the acquisition and retention of serial lists. *Journal of Verbal Learning and Verbal Behavior*, 1967, **6**, 721–728.

Postman, L., Turnage, T. W., & Silverstein, A. The running memory span for words. *Quarterly Journal of Experimental Psychology*, 1964, **16**, 81–89.

Saba, A. K. Inconsistencies in tests of the unit-sequence interference hypothesis in short-term memory. Unpublished master's thesis, University of Maryland, 1972.

Turnage, T. W. Unit-sequence interference in short-term memory. *Journal of Verbal Learning and Verbal Behavior*, 1967, **6**, 61–65.

Turnage, T. W. Proactive inhibition of initial items in short-term recall. *Psychonomic Science*, 1970, **20**, 236–237.

Turnage, T. W., & McCullough, T. A. Letter-sequence and unit-sequence effects during learning and retention. *Journal of Experimental Psychology*, 1968, **76**, 141–146.

Turnage, T. W., & McGinnies, E. M. A cross-cultural comparison of the effects of presentation mode and meaningfulness on short-term recall. *American Journal of Psychology*, in press.

Turnage, T. W., & Steinmetz, J. I. Unit-sequence interference and short-term recall. *American Journal of Psychology*, 1971, **84**, 112–122.

Underwood, B. J. Intralist similarity in verbal learning and retention. *Psychological Review*, 1954, **61**, 160–166.

Underwood, B. J. Degree of learning and measurement of forgetting. *Journal of Verbal Learning and Verbal Behavior*, 1964, **3**, 112–129.

Underwood, B. J., & Ekstrand, B. R. Word frequency and accumulative proactive inhibition. *Journal of Experimental Psychology*, 1967, **74**, 193–198.

Underwood, B. J., & Postman, L. Extra-experimental sources of interference in forgetting. *Psychological Review*, 1960, **67**, 73–95.

(Received July 10, 1972)

Journal of Experimental Psychology
1973, Vol. 98, No. 2, 335–338

ASYMMETRIC ATTRIBUTE INTERACTION IN CONCEPT IDENTIFICATION[1]

THOMAS D. WICKENS[2] AND BRIAN ZAX[3]

University of California, Los Angeles

The *S*s were run in a two-category affirmative concept identification task using geometric figure stimuli formed from six binary-valued attributes. An attempt was made to predict the rates of learning for one attribute from the amount that the learning rate of another attribute was increased when the first attribute was held constant. The relative attribute saliencies did not remain constant as attributes were added or deleted. There were interactions between the attributes that determine the saliencies, and these interactions sometimes affected one attribute more than another. The interactions may represent a source of difficulty in the evaluation of concept identification theory.

Recent studies of the two-category affirmative concept identification (CI) paradigm have supported a model of this task based on the testing of hypotheses or the checking of attributes (e.g., Bower & Trabasso, 1964; Wickens & Millward, 1971). It now appears that this class of models provides the most satisfactory explanation of *S*'s behavior. Many experiments that have attempted to probe these processes in more detail—for example, those using the redundant relevant cue paradigm (Trabasso & Bower, 1968)—have compared learning rates for several groups that differ in the size or nature of the stimulus set that *S* is allowed to see. To interpret these experiments it is necessary to assume that the changes in the stimulus set produce no major changes in other aspects of the solution process, in particular that the relative salience of the attributes remains roughly constant. It is the intent of this paper to demonstrate a potential failure of this assumption.

Central to all CI models is the process by which *S*s choose the relevant attribute, dimension, or hypothesis. An attribute (or set of attributes) is selected, then *S* checks its consistency with the observed pattern of stimulus classification. The probability that the correct attribute is chosen for examination is usually represented by a normalized weight. Thus attribute A_i has weight w_i and the probability that A_i is actually chosen for testing is

$$c_i = \frac{w_i}{W} \qquad [1]$$

where W is the total weight of all attributes that S could select, $W = \Sigma w_j$. The quantity c_i represents the probability that A_i is chosen for responding after an error, and will be referred to as the learning rate, since selection of A_i allows S to solve a problem based on that attribute. Although w_i cannot be determined, c_i can be estimated from the number of errors made while solving a problem based on A_i.[4]

If irrelevant attributes are added to the stimulus array, the denominator of Equation 1 is increased and the learning rate for A_i decreases. Similarly, holding an irrelevant attribute constant will increase c_i. In particular, Equation 1 implies that the increase in the learning rate for attribute A_i produced by holding attribute A_j constant can be predicted from the learning rate for A_j. In order for this analysis to be correct, the weight of A_i must be the same whether or not A_j is variable. If this is not the case, it suggests that weights estimated from one set of stimuli cannot

---

[1] This research was supported by National Institute of Mental Health Research Grant MH-19099-01 from the U. S. Public Health Service. The authors would like to thank Morton P. Friedman for comments on an earlier draft of this article.

[2] Requests for reprints should be sent to Thomas D. Wickens, Department of Psychology, University of California, Los Angeles, California 90024.

[3] Now at the Ohio State University.

[4] The mathematical form and derivation of the estimates and predictions used in this article can be obtained from the first author.

TABLE 1

LEARNING RATES, PARAMETER ESTIMATES, AND ERRORS OF ESTIMATION

Measure	Set DP				Set SF				Set PF			
	D	D-P	P[a]	P-D	S	S-F	F[a]	F-S	P[a]	P-F	F[a]	F-P
No. of solvers	15	15	14	14	12	15	14	15	14	15	14	15
No. of nonsolvers	0	0	1	1	3	0	1	0	1	0	1	0
Total no. of errors	89	25	180	154	282	117	101	155	180	138	101	96
$\hat{c}$	.144	.375	.072	.083	.040	.114	.121	.088	.072	.098	.121	.135
Estimated $\sigma_{\hat{c}}$	.034	.076	.018	.021	.011	.028	.029	.022	.018	.024	.029	.032
$\bar{c}$	.133		.615		−.368		.644		.107		.268	
Estimated $\sigma_{\bar{c}}$	.308		.121		.475		.131		.305		.259	
$\bar{c} - \hat{c}$ $(=\bar{c}^*)$	−.011		.543		−.408		.523		.035		.147	
Estimated $\sigma_{\bar{c}-\hat{c}}$	.311		.122		.475		.134		.305		.260	

[a] Group P is reported in both Sets DP and PF; Group F is reported in both Sets SF and PF.

necessarily be directly applied to another set and casts doubt on the usefulness of Equation 1 in the interpretation of restricted stimulus experiments.

METHOD

One hundred fifty male and female students from the introductory psychology course at University of California, Los Angeles served as *S*s as part of their course requirements. Each *S* was seated in a quiet room about 2 m. from a screen upon which the stimulus slides were projected. Instructions were read describing the task, then *S* was given 10 trials on a three-dimensional practice problem. The dimensions of this problem were determined by letters printed in upper- or lowercase and were chosen to provide minimum transfer to the experimental stimuli. After these 10 trials, the solution to the practice problem was explained to *S* and the experimental problem was run. The experimental stimuli were not described to *S* before the experiment.

The experimental stimuli were geometric figures formed from six independent binary dimensions: (*a*) the color of the figure (red or blue); (*b*) the size of the figure (large or small in a ratio of about 2.8 to 1 by area); (*c*) the shape of the figure (square or circle); (*d*) the position of the figure (left or right); (*e*) the position of a dot (above or below the figure); and (*f*) the color of the background (black or white). Trials were self-paced and commenced with the appearance of a randomly chosen slide. The *S* classified this stimulus verbally as "one" or "two," and then was told if he was right or wrong. Immediately thereafter the slide was removed and the next trial started. The only exception to the random selection of stimuli was on Trial 2 when the complement of the Trial 1 stimulus was presented to assure that *S* had seen all attributes. Criterion for solution was 10 consecutive correct responses. If *S* had not begun a criterion run by Trial 91, the session was terminated.

The *S*s were assigned randomly to 10 groups of 15 *S*s, which may logically be combined into three partially overlapping sets of 4 groups. Set DP contained four groups using the location of the dot and the position of the figure as relevant attributes. In Group P the position dimension was relevant and all six dimensions were varied; in Group P-D the position dimension was still relevant, but the dot dimension was held constant; Groups D and D-P were similar to P and P-D except that the dot was relevant and, in D-P, position was held constant. In Groups P-D and D-P the constant attribute took one of its values for half of the *S*s, the other for the remaining half. The second set of four groups, SF, used size and form as relevant or constant dimensions, and contained Groups S, F, S-F, and F-S. The remaining two groups, P-F and F-P, combined the form and position dimensions, and with the P and F groups from above, constituted the third, PF, set.

RESULTS AND DISCUSSION

Six of the 150 *S*s failed to solve the problem within 90 trials, 3 in Group S and 1 each in Groups F, P, and P-D. The total number of errors made by all *S*s is given in Table 1. The first trial for each *S* was a pure guess, contributed no information about the solution process, and has been ignored both in these counts and in the subsequent analysis. Maximum likelihood estimates of the learning rates can be obtained and are denoted in Table 1 by $\hat{c}$. Estimates of the asymptotic standard deviation of the estimates, $\sigma_{\hat{c}}$, are also given. It should be noted that these deviations are exact only when the number of *S*s is very large and provide only a rough idea of the accuracy of the estimates.

The result of restricting an irrelevant attribute was to increase the learning rate for the relevant attribute (in the one exception, Groups F and F-S, the change was within one standard error of the difference). A change in this direction is predicted by almost any model of CI. Equation 1 allows more precise predictions to be made about the relations among the learning rates for a set of four groups, however. If Equation 1 provides an adequate representation of attribute selection and if the weights remain constant over the different problems, it is possible to predict c_i from c_j and c_{i-j}. These predictions are given in Table 1, denoted by $\tilde{c}$. They are based on different data than is $\hat{c}$, but should be identical to it if the attribute weights are constant. The difference $\tilde{c} - \hat{c}$ is indeed less than its estimated standard error for four of the six estimates, but when c_P is estimated from c_D and c_{D-P} and when c_F is estimated from c_S and c_{S-F}, a large (greater than $3.5\ \sigma_{\tilde{c}-\hat{c}}$) discrepancy appears. The basic nature of this discrepancy is the same in both cases: When figure position or form is fixed, the learning rate of the dot position or the size increases to a greater extent than is consistent with the value of c_P or c_F alone. For these attributes a simple application of Equation 1 with constant weights is not satisfactory.

It may appear that a constant weight model can still be retained if the weight of the hypotheses that are eliminated when an attribute is held constant exceeds the weight of the relevant hypotheses for that attribute. In particular, one can postulate joint hypotheses such as "large squares or small circles are 1" that are present whenever both attributes (size and form) are variable, but absent whenever either is constant. Even with this modification constant weights are not supported. Normalized weights for joint hypotheses are estimated by the difference $\tilde{c}^* = \tilde{c} - \hat{c}$ (see Table 1). Sets DP and SF suggest that joint cues associated with the position and form dimensions have heavy weight, while the PF set suggests few joint cues associated with either. This contradiction shows that the analysis in terms of joint hypotheses is not satisfactory and that the size of $\tilde{c} - \hat{c}$ depends not only on the tested attribute, but also on the attribute that was held constant to obtain $\tilde{c}$.

To interpret these results, it is necessary to abandon the narrow mathematical formulation of Equation 1 and examine the specific dimensional relationships themselves. In the Set DP, the right or left placement of the entire figure was obviously bivalued, regardless of the position of the dot, so the weight assigned to it remained roughly constant. In contrast, when position was varied the dot could appear in four locations (upper left, lower left, etc.), yet when position was constant it could appear in only two. In the former case *S*'s initial reaction to the dot could have been as a four-valued attribute, giving it substantially greater complexity and giving its correct representation a low saliency.

The Set SF is very similar. One dimension, form, was clear regardless of the figure size, while the size dimension was partially obscured by the form. "Large" and "small" have no absolute meanings, making it difficult to compare the sizes of different forms; *S*s could not initially attend to size per se, but to large and small circles and large and small squares. For the form dimension, no such difficulties arise.

Finally, the PF set produced nothing approaching a significant inconsistency. This is plausible, for both form and position are easily identifiable attributes that are equally obvious regardless of the presence or absence of the other. They should be nearly independent of each other and satisfy a constant cue weight model.

In this experiment certain attributes, dot position and size, suffered a loss of saliency when other attributes, figure position and form, were allowed to vary. The obscured attributes were those that apparently took on a greater number of values when the obscuring attributes varied. This effect was undoubtedly accentuated by *S*'s relative unfamiliarity with the stimuli, but would probably be present unless the stimulus dimensions were well overlearned. These inconsistencies do not necessarily call Equation 1 into question. The overall process model of attribute selection can hold and Equation 1 be correct for a single set of stimuli, while the learning rate parameters for these stimuli do not transfer to another stimulus set. A similar result has already been noted for discrimination data (Shipley & Luce, 1964). The problem can be resolved only by a closer examination of attribute relationships. In some cases these relationships are probably the same as those described by Garner (Garner & Felfoldy, 1970) as integral or separable. However, Garner's stimuli were very simple, and only rarely did a suggestion of an asymmetric interaction appear. The stimuli in the present experiment were much more complicated and showed interaction patterns that were neither integral nor separable, but included aspects of both. In any case, it is clear that the relationships between the dimensions of the geometric stimuli were neither identical nor symmetric, and that if a full picture of the attribute sampling procedure is to be developed, a model of the dimensional interactions would need to be specified.

REFERENCES

Bower, G. H., & Trabasso, T. Concept identification. In R. C. Atkinson (Ed.), *Studies in mathematical psychology.* Stanford: Stanford University Press, 1964.

Garner, W. R., & Felfoldy, G. L. Integrality of stimulus dimensions in various types of information processing. *Cognitive Psychology*, 1970, 1, 225–241.

Shipley, E. F., & Luce, R. D. Discrimination among two- and three-element sets of weights. In R. C. Atkinson (Ed.), *Studies in mathematical psychology*. Stanford: Stanford University Press, 1964.

Trabasso, T., & Bower, G. H. *Attention in learning: Theory and research.* New York: Wiley, 1968.

Wickens, T. D., & Millward, R. B. Attribute elimination strategies for concept identification with practiced subjects. *Journal of Mathematical Psychology*, 1971, **8**, 453–480.

(Received July 10, 1972)

Journal of Experimental Psychology
1973, Vol. 98, No. 2, 339–343

EFFECT OF PATTERN IN DISPLAY BY LETTERS AND NUMERALS UPON ACQUISITION OF SERIAL LISTS OF NUMBERS[1]

ALLAN L. FINGERET[2] AND W. J. BROGDEN

University of Wisconsin—Madison

A list of 16 two-digit numbers was acquired by serial anticipation to a criterion of one errorless trial by 16 *S*s in each of seven conditions. Four experimental conditions provided display patterns by presenting half of the numbers as words (e.g., fifty seven) and half as numerals (e.g., 83). The sequential patterns of display were single, double, quadruple, and octuple alternation. Control conditions were all numerals, all words, and a random sequence of half numerals, half words. The group with single alternation showed significantly more time and errors to acquisition than the other experimental groups. Examination of the significant Groups × Serial Position interaction revealed differences in the form of the curve for each experimental group from that for the combined control group. The view was proposed that any effects of pattern of organization upon acquisition, including form of the serial position error curve, are primarily transfer effects from prior learning.

Lists of verbal items used in experiments on serial learning provide many opportunities for the serial arrangement of items. Such arrangements can be made on a random, haphazard basis or on some predictable or organized basis. Experiments that have compared the acquisition of these kinds of serial lists have almost universally found organized lists superior to unorganized lists. Guilford (1927) reported faster acquisition of serial lists of numerals with patterns of arithmetic progression and regression than with randomly arranged sequences of numerals. Asch, Hay, and Diamond (1960) compared acquisition for spatio-temporal displays of separate windows for each nonsense syllable of the list with the standard procedure of presenting each syllable at the same window. A triangular display of windows was found superior to a linear display and both were better than the standard display of a single window. Weingartner (1963) found that serial lists of word associates arranged to maintain intact associative structure were acquired at a faster rate than were lists of the same words arranged on a random basis. Miller (1965) compared the acquisition of single-category items and of mixed-category items and found faster acquisition for the latter kind of list. The mixed-category items were learned more rapidly when the items of each category were organized into groups than when they were intermingled at random, and organized lists with four categories were learned more rapidly than lists with two categories.

The present study was planned to investigate serial anticipation acquisition of lists of two-digit numbers as a function of systematic variation in the sequencing of item display by numerals (e.g., 37) or by letters (e.g., thirty seven). Sequencing of display may be used to provide patterns of single alternation (e.g., 37, eighty five, 29, fifty two, . . .); double alternation (e.g., 37, 85, twenty nine, fifty two, . . .); quadruple alternation (e.g., four numeral display items, then four letter display items, . . .); and octuple alternation (e.g., eight numeral display items, then eight letter display items) when list length is 16 and each display mode occurs with equal frequency. These conditions provide a progression over the patterns of SA, DA, QA, and OA that was hypothesized to reduce difficulty of acquisition. The progression involves an inverse relation between size of alternation unit and frequency of alternation. The ease of acquisition, however, was hypothe-

---

[1] This study was supported in part by a grant from the National Science Foundation.

[2] Requests for reprints should be sent to Allan L. Fingeret, Department of Psychology, Albion College, Albion, Michigan 49224.

sized to increase both when size of the alternation unit was increased and the frequency of alternation was decreased. It was further hypothesized that the pattern of display would be reflected in the form of the serial position error curve even in the absence of any effect of pattern upon ease of acquisition. All experimental patterns were expected to produce more efficient acquisition than control patterns that involved either a random sequence of display or consistent displays either by all letters or all numerals. The random display was hypothesized to be more difficult than either of the two consistent displays which were not expected to differ from each other, although the possibility was considered that on the basis of a greater frequency of general usage of numerals, the consistent numeral display might be superior to the consistent letter display. The finding of no difference between letter and numeral display, however, would demonstrate their stimulus equivalence and provide an important procedure for manipulating list patterns completely independent of any effects due to differences in characteristics of the stimulus used to produce patterns. The experiment to be reported was designed to test these hypotheses.

METHOD

Subjects. The 112 *Ss* were in classes of introductory psychology and received points toward their grades for participation in the experiment. They were assigned randomly in order of their appearance at the laboratory to the seven treatments of the experiment ($n = 16$/group). All *Ss* were tested by a single *E*.

Design. Organization of the serial lists of 16 two-digit numbers was provided by displaying half of the items by letters (words) and half by numerals in systematic sequence or pattern. The four patterns and designations for the experimental groups were single alternation (SA), double alternation (DA), quadruple alternation (QA), and octuple alternation (OA). With SA, the first item was displayed as a numeral (or word), the second item as a word (or numeral), the third item as a numeral (or word) and so on to the end of the list (e.g., 57, eighty three, 29, . . . forty two; or fifty seven, 83, twenty nine, . . . 42). For DA, the first two items were displayed as numerals (or words), the third and fourth items were words (or numerals), and so on to the end of the list (e.g., 57, 83, twenty nine, sixty five, . . . forty two; or fifty seven, eighty three, 29, 65, . . . 42). With QA, the first four items were displayed as numerals (or words), items five, six, seven, and eight were displayed as words (or numerals), items nine, ten, eleven, and twelve as numerals (or words), and the last four items as words (or numerals). For OA, the first eight items were displayed as numerals (or words) and the last eight items were displayed as words (or numerals). For the three control groups, all items of the serial list were presented as numerals (Group N), all items were presented as words (Group W), or half of the items were numerals and half were words and they were presented in a random sequence (Group R). For all conditions practice continued until attainment of the criterion of one errorless trial.

Materials. The 16-item lists of two-digit numbers were constructed so that any given digit occurred no more than twice in both the first and second ordinal positions. Neither zero nor double digits were used (e.g., 66). Two different lists were constructed, and one half of the *Ss* in each group learned each one of the two lists. For the groups of *Ss* in which both numerals and words were presented, counterbalancing of the locus of numerals and words occurred within each half of the *Ss*. For slides of the letter display, the first of the two words of each number was located above the second.

Procedure. Slides of the items for the serial lists were projected white on a blue background by a Kodak Carousel automatic projector on a "milk plastic" screen in an aperture of the experimental chamber at a 2-sec. rate. The *S* was seated inside the experimental chamber at a table such that the projection screen was approximately 75 cm. from his face with vertical height of 2.5 cm. of letters and digits at the screen. Slides of the letter A provided the anticipation cue for the first item. The intertrial interval was 10 sec. An intercom system between the experimental and control spaces permitted *E* to give instructions to *S* and to record the vocal anticipatory responses made by *S* in serial learning. Failures to respond and incorrect responses were recorded as errors.

The same instructions were given to *Ss* of all groups. No information was given about the SA, DA, QA, OA, or R display patterns. The serial anticipation method was described and termination of practice was noted to occur following one perfect trial. Preliminary practice was given on a list of two-letter words to ensure understanding of the task on the part of each *S*.

RESULTS

The experimental design provides for statistical analyses of errors and time per item to the acquisition criterion. Analyses of variance were used in each case and, when appropriate, Fisher's least significant difference test (LSD) of differences between means was made subsequent to

TABLE 1

Effect of Display Pattern upon Errors and Time per Item to the Acquistion Criterion

Measure	Display pattern							
	SA	DA	QA	OA	N	W	R	CC
Errors	345.2	265.0	275.8	250.9	270.8	320.2	295.3	295.4
Time per item (in sec.)	86.6	66.9	67.5	61.8	71.9	81.9	73.9	75.9

Note. SA = single alternation; DA = double alternation, QA = quadruple alternation; OA = octuple alternatipn; N = all numerals; W = all words; R = random sequence; and CC = combined control group.

analysis of variance. The 5% level was the criterion of significance for all statistical tests. All analyses of the error data provided an evaluation of serial position and its interaction with the treatment conditions.

The means for errors and time per item to the acquisition criterion are given for all groups in Table 1. In the analyses of both errors and time per item, groups was not a significant source of variation at the 5% level but was at the 10% level, *F* (6, 105) = 1.96 and 1.98, respectively. The three control group means differed relatively little and in spite of a significant Groups × Serial Position interaction with the error data, *F* (90, 1575) = 2.29, plots of the serial position error curves of the control groups showed virtually no differences. Separate analyses of the data for the three control groups showed no significant differences between groups for errors and time per item, *F* (2, 45) = 1.2 and .8, and no significant Groups × Serial Position interaction, *F* (30, 675) = 1.0. It was decided, therefore, to pool the data of Groups N, W, and R to form a single combined control group (CC) and conduct unequal-*n*, unweighted-mean analyses of variance for Group CC and the four experimental groups. Groups was a significant source of variation in the analyses of errors and time per item, *F* (4, 107) = 2.7 and 2.9, respectively, and the Groups × Serial Position interaction in the error analysis was significant, *F* (60, 1605) = 2.9. The LSD for errors was 62.8 and for time per item was 16.3 sec. The means for Group CC are given in the last column of Table 1. The results on the LSD tests for errors and time per item are the same. The mean for Group SA is larger and significantly different from the mean of each of the other three experimental groups (DA, QA, and OA). There are no other significant differences between the means.

In order to examine differences between the form of the serial position error curves independently of magnitude of error, the data for every *S* at each serial position were converted to percentage of his total error. The unequal-*n*, unweighted-mean analysis of variance of the percentage error data for Groups SA, DA, QA, OA, and CC showed the Groups × Serial Position interaction to be significant, *F* (60, 1605) = 2.9. The four panels of Figure 1 present a plot of the serial position percentage error curve of each of the experimental groups compared with the curve for Group CC. The LSD for any difference between a mean for Group CC and a mean for one of the experimental groups was 1.1%. For any difference between means within an experimental group or between experimental groups, the LSD was 1.4%.

The form of the serial position error curves for Groups SA and DA show substantial similarity. Beginning with Serial Position 6, a sawtooth single-alternation pattern is visible in the curve for Group SA to Serial Position 12 and in the curve for Group DA to Serial Position 14. The amplitude of the oscillation is small, and in one instance for Group SA and two instances for Group DA it exceeds the LSD. There are two serial positions (7 and 10) at which the differences between Groups SA and CC exceed the LSD and three serial positions (10, 15, and 16) at which the differences between Groups DA and CC exceed the LSD.

The patterns for Groups QA and OA

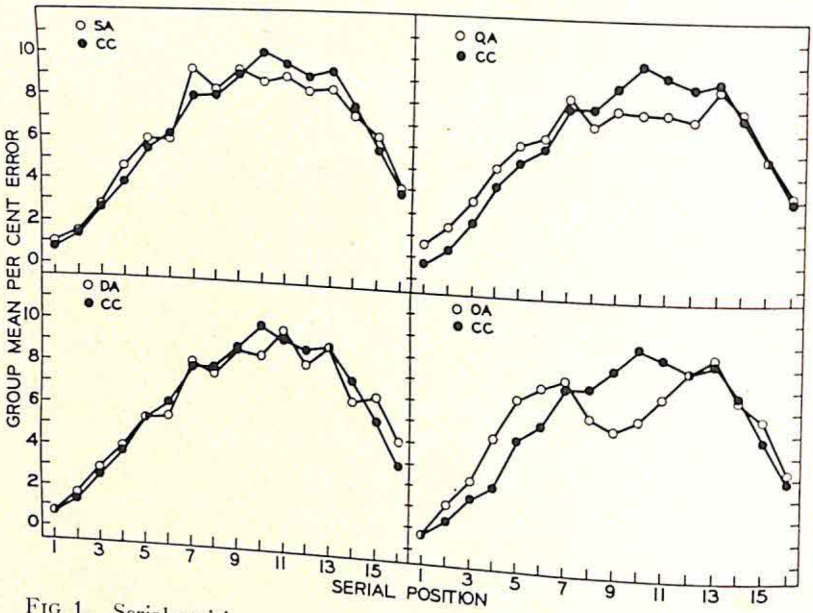


FIG. 1. Serial position error curves of the experimental groups (SA = single alternation; DA = double alternation; QA = quadruple alternation; and OA = octuple alternation) compared with the curve for the combined control group (CC).

show similar effects on the form of the serial position error curve. The early serial positions (1–7) for both experimental groups show a greater proportion of error than Group CC. For the middle serial positions (8 to 11 or 12) Groups QA and OA show a smaller proportion of error than Group CC. There are little or no differences between the experimental and control groups for the final serial positions (12 or 13 to 16). The differences between Groups QA and CC exceed the LSD at Serial Positions 3, 4, 10, 11, and 12, and the differences between Groups OA and CC exceed the LSD at Serial Positions 4, 5, 6, 9, 10, and 11.

DISCUSSION

The hypotheses about the effect of pattern of display for numbers by letters or numerals upon speed of acquisition were confirmed only to the extent that acquisition of the SA pattern was more difficult than any of the other organized patterns of display. The greater difficulty of acquisition for the SA pattern was possibly due to the high frequency of alternation and minimal block size of display. In spite of the limited effect of pattern of display upon ease of acquisition, the form of the serial position error curve for each of the experimental groups was found to differ significantly from the standard or control serial position curve. The SA and DA patterns and the QA and OA patterns were paired in showing similar effects on the form of the serial position curve. In both cases, the major effect occurred for the middle serial positions. In neither case did the effect of pattern override the effect of recency. There were no substantial differences between the curves for the experimental groups and the control group at the last three or four ordinal positions. The effect of primacy was dominant over any effect of the SA and DA patterns, since the curves for these experimental groups did not differ from the control group at the initial ordinal positions. The curves for the QA and OA patterns, however, showed a consistently greater magnitude of error at the early positions than did the control group. Whether this was an effect of pattern on magnitude of error at these positions or compensation for the reduction of error at the middle positions is difficult to say. The pattern effects, being most noticeable at points of change for both QA and OA, may have provided anchor points for part list learning of greater efficiency than whole list learning.

This latter hypothesis gains support from the results obtained by Fingeret and Brogden (1972) with DA and QA patterns of nouns and two-digit numbers in the combination list when the two part lists were pure lists of nouns and two-digit numbers. Part learning was more efficient than whole learning when the DA and QA patterns were composed of items from the part lists. Part learning of the DA and QA patterns with similar material in the combination list did not differ significantly from whole learning of the same pattern. The form of the serial position curves for the combination lists showed substantial effects of the DA and QA patterns that showed greater magnitude for specific than for nonspecific transfer from the part lists. The smallest magnitude of the pattern effects occurred with whole learning.

These results suggest that the effects of pattern are due in part to positive transfer from prior learning. Any pattern or principle that might be used for organization of a serial list to be learned by adult *S*s would have a substantial history of prior practice. Thus, the prior learning of a pattern or principle would always be confounded with the current effect such a variable would have upon the ease of acquisition of a serial list. Any conditions that contribute to increasing the magnitude of such positive transfer effects should contribute also to the effect of organization upon the difficulty of serial learning. Thus, when patterns or principles of organization are manipulated as variables, the prior learning and its positive transfer and the effect of pattern upon ease of serial learning are intermingled. The effects of primacy and recency may also be considered in the same manner. Positive transfer effects from prior serial learning are the basis for primacy and recency effects and these effects are prominent and large for serial learning. When there is an addition of some pattern or principle of organization to a serial list, demonstration of the effects of pattern are necessarily limited to the middle serial positions. Primacy and recency effects will interfere with and markedly reduce any pattern effects at the beginning and terminal ordinal positions. Even without the imposition of a specific pattern every serial list has a minimum pattern that is dependent upon positive transfer effects from previous serial learning that involves primacy and recency effects due to the initial and terminal ordinal positions. The acquisition of items in the middle of the list will occur at a slower rate than will those at the beginning and end of the list and will depend upon cues that are inherent in the items or upon cues that are idiosyncratic with each learner.

REFERENCES

ASCH, S. E., HAY, J., & DIAMOND, R. M. Perceptual organization in serial rote learning. *American Journal of Psychology*, 1960, **73**, 177–198.

FINGERET, A. L., & BROGDEN, W. J. Item arrangement effects on transfer and serial position errors in part–whole learning of different materials. *Journal of Experimental Psychology*, 1972, **93**, 249–255.

GUILFORD, J. P. The role of form in learning. *Journal of Experimental Psychology*, 1927, **10**, 415–423.

MILLER, G. R. Intrinsic organization of serial lists. *Psychological Reports*, 1965, **16**, 47–50.

WEINGARTNER, H. Associative structure and serial learning. *Journal of Verbal Learning and Verbal Behavior*, 1963, **2**, 476–479.

(Received July 17, 1972)

Journal of Experimental Psychology
1973, Vol. 98, No. 2, 344–349

MAZE LEARNING OF MATURE–YOUNG AND AGED RATS AS A FUNCTION OF DISTRIBUTION OF PRACTICE[1]

CHARLES L. GOODRICK[2]
Gerontology Research Center, National Institute of Child Health and Human Development, Baltimore, Maryland

Mature-young and aged rats were tested in a 14-unit multiple T-maze under the conditions of distributed or moderately massed practice. Mature-young rats learned the maze more quickly than aged rats under both practice conditions, but mature-young rats learned most efficiently with distributed practice, while aged rats learned most efficiently with massed practice. In a second experiment, aged rats which failed to learn the maze were given distributed, massed, or highly massed practice. The massed practice conditions resulted in fast and efficient reduction in error scores. These data are best understood in terms of a deficit in short-term memory for aged rats.

Kimble (1961) summarized evidence concerning the effect of massed practice upon learning: "The typical effect of the crowding of practice trials in time . . . is to diminish performance to a degree which depends upon the degree of massing . . . [p. 125]." Although Kimble referred to experiments involving human learning, many early experiments with rats exist, most of which are congruent with this summary (Munn, 1950, pp. 319–323). An exception to this rule apparently occurred in an experiment involving mature-young and aged albino rats (Goodrick, 1968). Aged rats which had failed to reach a criterion of learning a complex T-maze were given highly massed practice and quickly learned the maze.

The present two experiments were designed to investigate this phenomenon further. The first experiment determined differences in rate of learning for groups of mature-young and aged rats tested under conditions of distributed or moderately massed practice. A second experiment utilized groups of aged rats in order to examine the effect of degree of massed practice upon rate of learning and efficiency of learning for slow-learning aged rats which had failed to reach a criterion during distributed practice tests.

EXPERIMENT I

Method

Subjects. Sixty-four male Wistar rats were used. At the time of testing 32 were 6 mo. old, and 32 were 26 mo. old. All of the rats were born, weaned, and then maintained in group cages, four or five per cage, within the Gerontology Research Center animal colony, until removal for the present experiment. The ages were selected to represent fully grown young rats and rats of advanced old age. The point of 50% mortality for these male Wistar rats is about 26 mo.; i.e., 50% of a given group die by the age of 26 mo., while the maximum longevity is about 32 mo. It is standard procedure in the laboratory to test an excess of aged rats which are randomly selected as replacements in the case of the death of an experimental *S*.

Apparatus. A modified Stone 14-unit spatial T-maze was used which was illustrated in a previous report (Goodrick, 1968, p. 299). The reliability of this maze was between .90 and 1.00, according to four different methods of determining reliability (Stone & Nyswander, 1927). The maze floor and sides (37.5 cm. high) were constructed of clear plastic, and maze alleys were 12.5 cm. wide to allow passage of the large adult male rats. Sliding plastic panels were used at the start chamber and Choice Points 3, 6, 9, and 12 to limit back tracing, with dummy panels opposite all true doors. A straight runway, used in preliminary training, was 100 × 12.5 × 45 cm. and was also constructed of clear plastic.

Procedure. Upon removal from the animal colony, all experimental animals were isolated in individual cages and given food and water ad lib for 3 days. The rats were then placed on a reduced food ration for the remainder of the experiment. Weights were lowered to 80% of starting weight for mature-young rats and 75% of starting weight for aged rats.

---

[1] The author would like to acknowledge the aid of John Freeman, who assisted in the collection of the data within this report.

[2] Requests for reprints should be sent to Charles L. Goodrick, Gerontology Research Center, Baltimore City Hospitals, Baltimore, Maryland 21224.

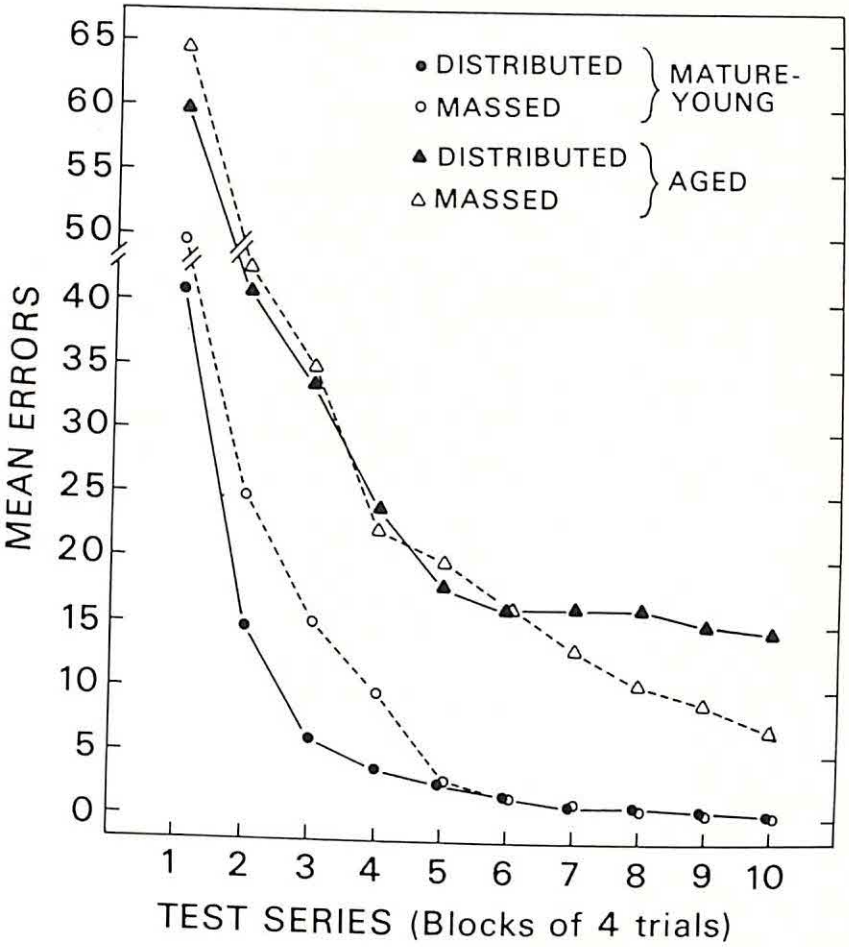


FIG. 1. Mean errors as a function of age, practice condition, and series (blocks of four trials).

Such differential deprivation favoring the aged animals was suggested by Stone (1929, pp. 123–124) and Jerome (1959). However, because mature-young rats normally increase body weight when isolated, the motivation of the two groups was probably equivalent (Stone, 1929, p. 104). The rats were always tested between 9 A.M. and 1 P.M. and fed their daily food ration between 2 P.M. and 4 P.M. After 15 days of food deprivation, the rats were adapted to drinking the reward solution of condensed milk and sucrose (sucrose/milk = 10 gm/100 ml) within the home cage. This solution was placed in a depression on heavy culture slides (used also in the runway and T-maze). Then all rats were gentled and given 10 preliminary training trials (2 per day) in the straight runway with 1 ml. of reward solution at the end of the runway. Following this procedure, testing in the 14-unit T-maze began. One half of each age group was tested with one trial per day, while the other half was tested with four trials per day (a series), with 30 min. between trials. All rats received a total of 40 trials and were rewarded in the goal area with 1 ml. of reward solution.

During T-maze tests, each rat was initially placed in the start area; then the panel was raised allowing access to the maze. A timer was started when the rat was fully outside the start chamber and stopped when the food reward was reached. If the forepaws were placed in an area which diverged from the true path to the goal, an error was counted. Forward errors and retracing errors were all counted, but during retracing, multiple errors were not assigned as different maze segments were traversed (Goodrick, 1968, Figure 2, p. 300; Munn, 1950, pp. 255–266, especially Figure 111, p. 260). The path through the maze was outlined on mimeographed sheets of the maze design during every trial.

Results

In both Experiments I and II, error scores were used to determine group differences in rate of learning. The utilization of the error score is especially important when aged rats are tested because time scores may be related to age differences in locomotor ability rather than age differences

TABLE 1

Analysis of Errors as a Function of Practice Condition and Age

Age	Practice condition	Trials 1–20			Trials 21–40		
		a	*b*	*c*	*a*	*b*	*c*
Young	Distributed	5.4	.9	14.3	.4	0.0	0.0
	Massed	6.5	1.4	19.8	.4	0.0	0.0
Aged	Distributed	7.1	3.8	50.5	3.8	2.5	64.9
	Massed	6.9	3.4	49.1	2.9	1.3	34.2

Note. (*a*) Mean number of culs-de-sac where six or more errors occurred; (*b*) mean number of perseverative errors; and (*c*) mean percentage of perseverative errors.

in learning. Mean errors for four-trial blocks were scores used in a 2 × 2 × 10 (Age × Practice Condition × Series) repeated-measures analysis of variance design. The mean scores are given in Figure 1.

Mature-young rats learned the maze more quickly than aged rats for both practice conditions, *F* (1, 60) = 109.37, $p < .001$. In addition, error scores were significantly reduced over series, and the Age × Series interaction was highly significant, *F* (9, 540) = 29.24, $p < .001$, because age differences were reduced as a function of series. The Practice Condition (distributed vs. massed) × Series interaction was also highly significant, *F* (9, 540) = 23.97, $p < .001$, because when age was not considered, massed practice groups initially had higher mean error scores than distributed practice groups, but during the last four series, massed practice groups had lower mean scores than distributed practice groups. However, the second order Age × Practice Condition × Series interaction was also significant, *F* (9, 540) = 5.50, $p < .001$. This interaction reached a significant level because mature-young rats reduced errors more quickly under the condition of distributed practice than massed practice, while the aged rats reached a lower level of errors under the condition of massed practice than distributed practice. Moreover, significant differences as a function of practice condition occurred early in training for mature-young rats (Figure 1, Series 2, 3, and 4), but occurred late in training for aged rats (Figure 1, Series 8, 9, and 10). Although age differences are apparent on the first series of four trials, these differences were not significant for Trial 1 (mature-young mean = 15.1; aged mean = 16.5).

Because the exact path of each rat was obtained during training, it was possible to make a finer analysis of the learning process. A detailed analysis was made of the number of repetitive or perseverative errors of mature-young and aged rats during early (Trials 1–20) and late (Trials 21–40) stages of learning. Perseverative errors were defined as six or more consecutive errors at the same cul-de-sac. Table 1 gives the mean number of culs-de-sac where six or more errors were obtained, the mean number of culs-de-sac where perseverative responses were obtained, and the mean percentage of perseverative responses. The probability of obtaining six consecutive errors in a two-choice situation is .014. The probability of obtaining this sequence may be slightly greater for the aged groups than for the mature-young groups because of higher error scores at all choice points, but this factor is offset because the perseverative error sequences of the aged rats were normally greater than the minimum of six consecutive errors. Mature-young rats had a lower percentage of perseverative responses than aged rats; for mature-young rats vs. aged rats, all comparisons, *t* (30) > 2.04, $p < .05$. Due to the large number of scores of zero, the mature-young groups were not included in an analysis of variance for the percentage scores. The analysis was a 2 × 2 (Practice Condition × Trials) design, using just the percentage scores of

TABLE 2

Errors at 33 Perseverative Choice Points as a Function of Series and Trials for Aged Rats Given Massed Practice

Trial	Series						
	A	B	C	D	E	F	G
1	33	31	25	18	7	5	0
2	33	23	13	10	5	1	0
3	33	17	12	5	0	1	0
4	33	5	4	2	0	0	0
Total perseverative choice points	33	33	26	21	8	5	0

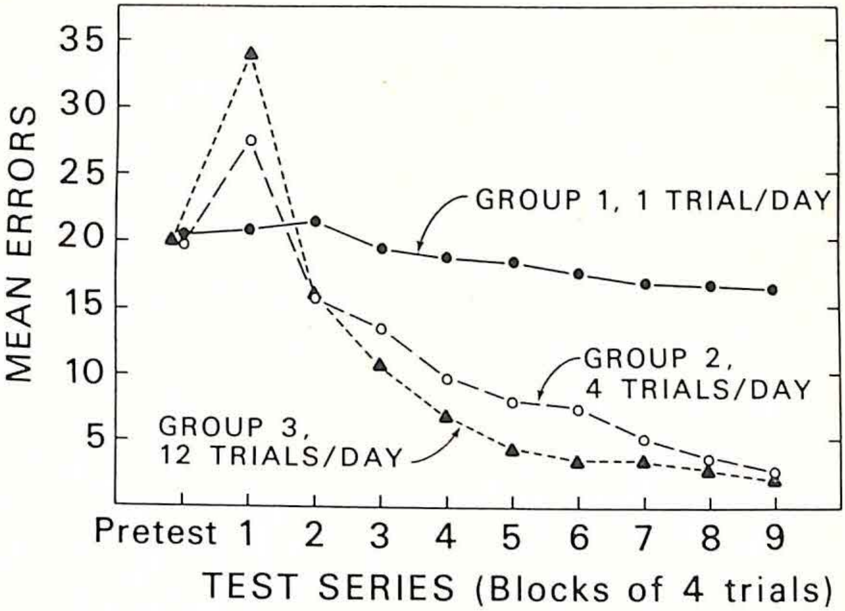


FIG. 2. Mean errors for slow-learning aged rats as a function of test series and distribution of practice.

the aged rats. The aged distributed practice groups had a significantly higher percentage of perseverative errors than the aged massed practice groups, F (1, 30) = 10.04, $p < .01$. This was due to an increase in the percentage of such errors over trials for the aged distributed practice groups, while a decrease in the percentage of such errors occurred over trials for the aged massed practice groups, Practice Condition × Trials, F (1, 30) = 12.24, $p < .01$. Although the group percentage differences were not statistically significant on Trials 1–20, group differences were highly significant for Trials 21–40, t (30) = 6.75, $p < .01$. Also, for the massed practice group, errors were plotted for choice points which were perseverative during early trials, but not during late trials (see Table 2). In Table 2, Series A shows errors made during the first perseverative series of four trials for 33 culs-de-sac. For Series A, errors were made on every trial, because perseverative errors are defined in terms of consecutive errors at specific culs-de-sac. Series B represents the first series in which less than four errors were made following at least six consecutive errors. These error plots indicated that errors were made more frequently on the initial trials of a daily series, rather than randomly over the daily series.

EXPERIMENT II

Method

Subjects. Twenty-four male Wistar rats, 26–28 mo. old, were *Ss*.

Apparatus. The apparatus was the same as that used in Experiment I.

Procedure. The preliminary motivational and adaptation procedures were the same as in Experiment I. The *Ss* were part of a larger group which were given 20 daily training trials. The rats utilized in the present experiment were classified as slow-learning rats if they failed to reach a criterion of four total errors or less on Trials 17–20. The slow-learning rats were then matched on the basis of total errors on Trials 17–20 and divided into three groups with eight rats in each group. All groups then received 36 additional trials in the maze; Group 1 received 1 trial per day, Group 2 received 4 trials per day, and Group 3 received 12 trials per day with a 60-min. interval between 4-trial sequences. Due to the large number of trials for Group 3, the between trial interval during four-trial sequences for Groups 2 and 3 was 10 min.

Results

Mean error scores for blocks of four trials were used in a 3 × 9 (Practice Condition × Series) analysis of variance design. These mean scores are given in Figure 2. Because the aged rats were matched on the basis of preliminary distributed practice scores, pretest means were very similar for the three groups. Following preliminary practice, group error differences were highly significant, F (2, 21) $= 10.23$, $p < .001$. On Series 1 of massed practice, massed practice Groups 2 and 3 increased error scores compared with rats of Group 1, which continued distributed practice training, t (22) $= 2.80$, $p < .02$. Both massed practice groups then reduced error scores at a much faster rate than the distributed practice group, Groups × Trials, F (16, 168) $= 7.43$, $p < .001$. Both massed practice groups obtained significantly fewer errors than the distributed practice group on test Series 7, 8, and 9, all ts (14) > 2.15, $p < .05$. Although more highly massed practice (12 trials per day) resulted in faster error reduction than less highly massed practice (4 trials per day), the differences between these groups were not statistically significant.

General Discussion

In the initial experiment, aged rats were found to learn the maze problem more slowly than mature-young rats under both practice conditions. This was consistent with results of an earlier experiment (Goodrick, 1968) which found that mature-young rats learned this maze pattern more quickly than aged rats. Both Birren (1964) and Botwinick (1967), in comprehensive textbooks on the psychology of aging, concluded that age differences in learning had not yet been successfully demonstrated for lower animals. However, the majority of research studies (e.g., Birren, 1962; Botwinick, Brinley, & Robbin, 1962) used simple learning tasks, which may have been of insufficient complexity, an important methodological factor emphasized by Munn (1950, p. 258). A more recent study of learning differences between mature-young and aged rats utilizing tasks which varied in complexity found reliable age differences in learning on only the most complex task, a 14-unit T-maze (Goodrick, 1972).

The finding of primary importance in the present experiment is that massed practice is a necessary condition for efficient learning of many aged rats. This is consistent with other results (Goodrick, 1968) which found that slow-learning aged rats were able to master a complex maze problem using highly massed practice trials (12 per day). Earlier suggestions of Stone (1929, pp. 117–124) and Verzar-McDougall (1957) were that massed practice (more than one trial per day) would tend to favor young more than aged rats, while distributed practice was the optimal practice condition for both young and aged rats. The two conditions involving massed practice of aged rats after training with distributed practice resulted in an initial increase in errors when compared with scores obtained by aged rats continuing distributed practice trials. This factor, which was also apparent in the earlier study (Goodrick, 1968), could explain the observations of Stone and Verzar-McDougall. For aged rats, moderately massed practice resulted in significantly fewer errors on the last trials than for aged rats tested with distributed practice (Experiment I). Mature-young rats, in contrast, learned more efficiently under the condition of distributed than massed practice. Mayer and Stone (1931) also found that one trial per day was optimal for mature-young rats tested in a maze similar to the one used in the present study. The massed practice conditions of Experiment II resulted in a rapid decrease in errors for these aged groups when compared with error scores of the aged group given continued distributed practice.

Welford (1958) had suggested that human age deficits in learning and performance may be the direct result of age differences in short-term retention. This analysis of the learning process was based on data obtained by Kay (1951) and Kirchner (1958) for learning performance of mature-young and aged human *S*s during a key-pressing task. The short-term memory retention mechanism is assumed to store information rapidly, but impermanently, after which a more durable trace must be established for long-term retention, which may be capable of remaining in storage unimpaired by the activities of the organism (Welford, 1958, pp. 224–256). The present data lend support to the hypothesis that learning deficits occur in aged organisms due to age differences in short-term retention because the aged groups given massed practice reduced error scores more efficiently than aged groups given distributed practice.

The complex problem used allowed a fine degree of analysis of the learning process. Aged rats tended to make repetitive errors at the same choice points on successive trials to a greater degree than mature-young rats. The percentage of perseverative errors increased with trials for aged rats given distributed practice, although the absolute number of errors, including perseverative errors, decreased. These rats tended to take the same path through the maze on the later trials and were unable to reduce the percentage of perseverative errors as were the aged rats given massed practice. The reduction of repetitive errors by aged rats given massed practice of four trials per day (Experiment I) occurred late in a four-trial daily series. This mode of reduction of repetitive errors is consistent with the hypothesis that such errors were the result of a short-term memory deficit.

REFERENCES

Birren, J. Age differences in learning a two-choice water maze by rats. *Journal of Gerontology*, 1962, **17**, 207–213.

Birren, J. *The psychology of aging.* Englewood Cliffs, N. J.: Prentice Hall, 1964.

Botwinick, J. *Cognitive processes in maturity and old age.* New York: Springer, 1967.

Botwinick, J., Brinley, J. F., & Robbin, J. S. Learning a position discrimination and position reversals by Sprague-Dawley rats of different ages. *Journal of Gerontology*, 1962, **17**, 315–319.

Goodrick, C. Learning, retention, and extinction of a complex maze habit for mature-young and senescent Wistar albino rats. *Journal of Gerontology*, 1968, **23**, 298–304.

Goodrick, C. Learning by mature-young and aged Wistar albino rats as a function of test complexity. *Journal of Gerontology*, 1972, **27**, 353–357.

Jerome, E. Age and learning-experimental studies. In J. Birren (Ed.), *Handbook of aging and the individual.* Chicago: University of Chicago Press, 1959.

Kay, H. Learning of a serial task by different age groups. *Quarterly Journal of Experimental Psychology*, 1951, **3**, 166–183.

Kimble, G. *Hilgard and Marquis' conditioning and learning.* New York: Appleton-Century-Crofts, 1961.

Kirchner, W. Age differences in short term retention of rapidly changing information. *Journal of Experimental Psychology*, 1958, **55**, 532–538.

Mayer, B., & Stone, C. The relative efficiency of distributed and massed practice in maze learning by young and adult albino rats. *Journal of Genetic Psychology*, 1931, **39**, 28–49.

Munn, N. *Handbook of psychological research on the rat.* Boston: Houghton Mifflin, 1950.

Stone, C. The age factor in animal learning: I. Rats in the problem box and the maze. *Genetic Psychology Monographs*, 1929, **5**, 1–130.

Stone, C., & Nyswander, D. The reliability of rat learning scores from the multiple-T maze as determined by four different methods. *Journal of Genetic Psychology*, 1927, **34**, 497–524.

Verzar-McDougall, E. Studies in learning and memory in ageing rats. *Gerontologia*, 1957, **1**, 65–85.

Welford, A. *Ageing and human skill.* Oxford: Oxford University Press, 1958.

(Received July 17, 1972)

Journal of Experimental Psychology
1973, Vol. 98, No. 2, 350–361

SOME STUDIES OF WORD ABBREVIATION BEHAVIOR[1]

MILTON H. HODGE[2] AND FLORRIE M. PENNINGTON

University of Georgia

Experiment I demonstrated that the percentage of letters used in constructing abbreviations of common words decreased with word length (4–5, 6–7, 8–9 letters) and Thorndike-Lorge (T-L) frequency (1–24, 25–49, A, AA). The contraction principle (omitting interior letters) was preferred with short words and truncation (omitting terminal letters) with long words. Number of different abbreivations increased with word length, but not T-L frequency. Experiment II asked *S*s to construct abbreviations by systematic use of contraction, truncation, or both. The results were virtually the same as before and showed that *S*s can use the abbreviation rules in a consistent fashion. Experiment III required *S*s to reconstitute the original words from the abbreviations produced in Experiment I. Although approximately 67% were reconstructed correctly, word reconstruction varied greatly with T-L frequency, word length, and sex of *S*. The experiments were related to intraword constraint and coding processes.

In written communications, word abbreviations play an important role in the general efficiency of language usage. They save time and require less effort than words, especially when particular words are used over and over as with the days of the week and the months of the year. Abbreviations are also essential in computer systems, certain signs, and various information displays. They occupy less space, and thus are less expensive than words.

Yet, despite their clear importance in many situations, previous concern with abbreviations has been limited to unproven prescriptions for their construction (Barrett & Grems, 1960; Bourne & Ford, 1961; Luhn, 1956). The prescriptions range from simple rules such as "omit all vowels" to the comprehensive abbreviation practices of scientific journals (e.g., *Chemical Abstracts*) and large industrial corporations (e.g., American Telephone and Telegraph Co.).[3] Although compelling arguments can be devised for each rule, apparently none has been subjected to empirical test. There seems to be no behavioral evidence to support a choice among the many systems proposed for the construction of abbreviations.

An analysis of existing practices suggested that most abbreviations make use of the principles of truncation, contraction, or a combination of the two. In *truncation*, one or more contiguous terminal letters of the word are omitted. Some examples are: pres. (president), Oct. (October), and min. (minute). An extreme form of truncation is exemplified by the single-letter abbreviations used with units of measurement (V for volts, F for Fahrenheit, etc.). In *contraction*, one or more interior letters are removed and the remaining letters pushed together. Some examples are: bldg. (building), amt. (amount), and rd. (road). The contraction principle does not require the deleted letters to occupy contiguous positions in the original word; any letters other than the initial and the terminal letters may be removed. Finally, some abbreviations embody both truncation and contraction. The simultaneous use of both rules seems fairly rare, but some familiar examples are: appx. (approximate), pvt. (private), and chm. (chairman). To illustrate the principles further, suppose an eight-letter word is to be abbreviated. If Letters 6, 7, and 8 are deleted, the abbreviation would illustrate truncation. If, however, Letters 3 and 5 are removed, the abbreviation would be an example of contraction. Omitting Letters 5, 7, and 8 would illustrate both truncation and contraction. The joint category is somewhat ambiguous in that removal of some letters (e.g., 7) could be governed by truncation or contraction, but there is no ambiguity when a word is abbreviated exclusively by truncation or contraction. The joint category is needed for those cases which cannot be clearly classified as truncation or contraction. Fortunately, as the present data show, very few abbreviations fit the joint category.

---

[1] This research was supported by a research contract between Bell Telephone Laboratories and the University of Georgia. The helpful comments of Karl P. Hunt, Marsie M. Field, and S. David Leonard are gratefully acknowledged.

[2] Requests for reprints should be sent to Milton H. Hodge, Department of Psychology, University of Georgia, Athens, Georgia 30601.

[3] The procedures employed by scientific journals are public knowledge and can be obtained from the editors. The practices of industrial firms are usually treated as proprietary information and are not readily available.

The present report describes three experiments designed to explore the usefulness of the three principles of abbreviation and to examine the effects of certain stimulus variables. In particular, the studies sought to determine the relative use of each principle in the actual construction of abbreviations in a free-choice situation (Experiment I) and to evaluate the consistency with which *S*s construct abbreviations when they deliberately use one of the principles (Experiment II). Despite the lack of previous research directly relevant to the present work, the investigations of language behavior by Miller and others (in Garner, 1962) suggested that the length of a word (number of letters) and its frequency of occurrence in the English language (by Thorndike-Lorge, 1944, norms) might influence the construction of abbreviations. Specific predictions were not attempted, but it seemed obvious that *S*s might treat the abbreviation of short words as trivial and unnecessary. On the other hand, abbreviations of long words represent meaningful and obvious economies in language usage, and hence might be assigned greater value. In terms of the frequency variable, it was hypothesized that the abbreviations of high-frequency words would show less variability than those of low-frequency words because *S*s would probably have had considerable experience in constructing abbreviations of high-frequency words. Both variables were manipulated in each experiment. Lastly, Experiment III was designed to assess the accuracy with which independent (new) *S*s could reconstitute the stimulus words from the abbreviations devised in Experiment I. The third study provided a test of the validity of the abbreviations. It should be emphasized that no attempt was made to obtain normative information about word abbreviation behavior. When more is known about the factors which influence abbreviations, normative studies might be useful.

EXPERIMENT I

The initial study sought to examine the effects of four experimental variables (word length, Thorndike-Lorge frequency, instructional set, and sex) to determine the frequency of use of the three principles and to evaluate several response measures. Early in the investigation, it was necessary to decide whether *S*s should be asked to construct abbreviations with a fixed number of letters or whether they should be given a free choice in the length of the abbreviation. Two considerations dictated the second approach. First, a free-choice situation seemed essential to test the variability prediction made for the frequency variable. Second, the number of letters in an abbreviation was more susceptible to quantification than other possible response measures, and thus more likely to differentiate among the treatment conditions. The instructional variable was also introduced to solve a methodological problem. Without specific guidance, it was speculated that some *S*s might generate meaningless and highly idiosyncratic abbreviations. Therefore, *S*s were instructed to provide for each word a *personal* abbreviation that they might use in their own private writing and a *general* abbreviation that could be understood and used by other persons.

Method

Subjects. The *S*s were 24 male and 24 female undergraduate students enrolled in the introductory psychology course at the University of Georgia. They participated to fulfill a course requirement.

Design. Independent *S*s (eight males and eight females) were randomly assigned to three word-length conditions, designated as short (4- and 5-letter words), medium (6- and 7-letter words), and

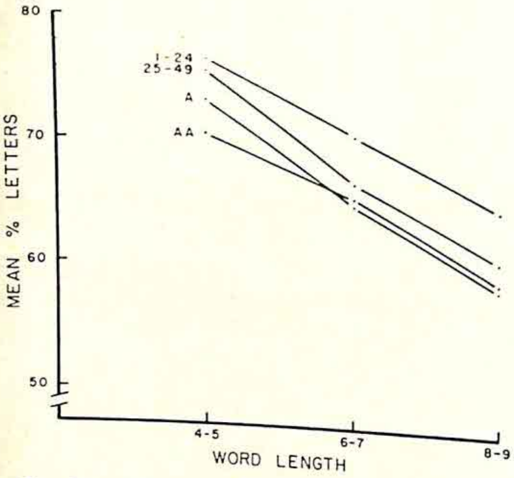


FIG. 1. Mean percentage of letters used in general abbreviations as a function of word length and Thorndike-Lorge frequency (Experiment I).

long (8- and 9-letter words). Four levels of Thorndike-Lorge (T-L) frequency were administered to all *S*s in each word-length condition. The T-L frequencies were 1–24 (per million words), 25–49, A, and AA. All *S*s in each condition were asked to construct a personal and a general abbreviation for each word. Sex and word length were treated as between-*S*s factors and T-L frequency and instructional set as within-*S*s variables.

Stimulus materials. For each combination of word length (4, 5, 6, 7, 8, and 9 letters) and T-L frequency (1–24, 25–49, A, AA), nine nouns were randomly selected from the Thorndike-Lorge word lists. Numbers, colors, and all proper nouns, including days, weeks, months, etc., were excluded from the lists. All words were verified for spelling and grammatical classification in Webster's Collegiate Dictionary (1966). The short-, medium-, and long-word-length conditions were formed by combining the 4- and 5-, 6- and 7-, and 8- and 9-letter words, thereby producing a list of 72 words for each word-length condition, with 18 words for each of the four T-L levels.[4] The order of presentation of the 72 words was then randomized, with the restriction that no more than two words from the same T-L level could occur in succession. A booklet consisting of instructions and nine pages of stimulus words was prepared for each word-length condition. Each page contained eight stimulus words and corresponding response spaces for the personal and general abbreviations.

Procedure. The *S*s, in groups of 4–8, were seated around a table and provided with a booklet. The instructions, read aloud by *E* while *S*s followed a written version, asked *S*s to construct a personal and general abbreviation for each word and to write them in the appropriate places in the booklets. They were told they could devise the abbreviations in any way they liked and that there was no limit on the number of letters to be used. Each *S* was asked to cover the preceding words and abbreviations with a blank sheet of paper as he moved down a page. An example of a personal and a general abbreviation was provided in the instructions, but no mention was made of the types of abbreviations or the rules for constructing them. Any questions were answered by repeating the instructions as nearly as possible. The *S*s were told that they had unlimited time, but to work as fast as seemed reasonable. Completion of the task required 15–30 min.

Results

Three analyses were undertaken, each based on a different response measure. The results are presented below.

Number of letters. In all conditions of the experiment, the number of letters in each abbreviation was determined and then converted to a percentage of the total letters of the word to be abbreviated. The mean percentage of letters employed in the general abbreviations is shown in Figure 1 as a function of word length and T-L frequency. Data for the personal abbreviations are not presented because they parallel those of the general abbreviations. The main difference is that about 5% fewer letters were used in the personal abbreviations. Several results are evident in Figure 1: (*a*) The percentage of letters declined systematically with increasing word length, although the actual number of letters showed a moderate increase over the three word lengths ($\bar{X} = 3.31$, 4.23, 5.17 for 4–5, 6–7, 8–9, respectively). (*b*) A larger percentage of letters was used in the low-frequency (1–24 and 25–49) than in the high-frequency (A and AA) words (the mean actual number of letters for each T-L frequency may be estimated from Figure 1 by multiplying the plotted values by 4.5, 6.5, or 8.5). Separate analyses of variance of the personal and general abbreviations indicated that significant differences were produced by word length: personal, $F\ (2, 42) = 8.08$, $p < .005$, general, $F\ (2, 42) = 11.18$, $p < .001$; T-L frequency: personal, $F\ (3, 126) = 33.52$, $p < .001$, general, $F\ (3, 126) = 48.22$, $p < .001$; and the interaction between these variables: personal, $F\ (6, 126) = 2.44$, $p < .05$,

---

[4] A list of the stimulus words may be obtained from the first author.

TABLE 1

FREQUENCY OF USE OF ABBREVIATION PRINCIPLES AS A FUNCTION OF WORD LENGTH, THORNDIKE-LORGE FREQUENCY, AND SEX

Abbreviation principle	Thorndike-Lorge frequency									
	Male					Female				
	1–24	25–49	A	AA	Total	1–24	25–49	A	AA	Total
	Length 4–5									
T	1	1	2	0	4	2	4	3	6	15
C	14	17	16	13	60	16	14	14	11	55
TC	3	0	0	5	8	0	0	1	1	2
	Length 6–7									
T	4	7	7	9	27	8	9	9	10	36
C	13	10	8	9	40	10	8	7	8	33
TC	1	1	3	0	5	0	1	2	0	3
	Length 8–9									
T	1	7	13	11	32	8	10	16	14	48
C	15	10	4	6	35	8	8	2	3	21
TC	2	1	1	1	5	2	0	0	1	3

Note. T = truncation; C = contraction; and TC = truncation-contraction.

general, $F(3, 126) = 3.26$, $p < .01$. Figure 1 suggests that the interaction is the result of a convergence of the effects of the 25–49, A, and AA T-L frequencies for the longer words. There is little difference in the percentage of letters used to abbreviate the 8–9 letter words with these three T-L values. Furthermore, as a group, these words were abbreviated with 5% fewer letters than the 1–24 words.

Frequency of abbreviation principles. Table 1 shows the number of words abbreviated with the three principles as a function of word length, T-L frequency, and sex.[5] Again, since the general and personal abbreviations closely parallel each other, only the data of the former will be described. Over all conditions, 244 (57%) abbreviations were based on the contraction principle, 162 (38%) involved truncation, and 26 (6%) followed both principles. However, the data also show that contraction was universally preferred only by male *S*s and mainly for the 4–5 and 6–7 letter words. In contrast, female *S*s preferred truncation, particularly with the 8–9 letter words. These effects were confirmed by chi-square analyses of the 3 × 3 contingency tables for word length and abbreviation principle (T-L frequency was collapsed). The tests showed significant differences among the frequencies for the males, $\chi^2(4) = 39.99$, $p < .01$, and the females, $\chi^2(4) = 33.48$, $p < .01$.

Other chi-square tests (two at each word length) indicated that T-L frequency was important only for the abbreviations of the 8–9 letter words. Significant differences were found in the 4 × 3 (T-L Frequency × Abbreviation Principle) contingency tables of both the males, $\chi^2(6) = 19.10$, $p < .01$, and the females, $\chi^2(6) = 12.81$, $p < .05$. Table 1 shows that truncation was used more often than contraction with the longer A and AA words. The reverse

---

[5] Since each word was abbreviated by eight *S*s (eight males or eight females in Table 1), classification of a word as truncation, contraction, or both was based on the principle used most often by the eight *S*s. For example, an entry of 14 indicates that 14 words were abbreviated with a particular principle by a majority of the *S*s.

TABLE 2

MEDIAN NUMBER OF DIFFERENT GENERAL ABBREVIATIONS AS A FUNCTION OF WORD LENGTH, THORNDIKE-LORGE FREQUENCY, AND SEX

Word length	Thorndike-Lorge frequency									
	Male					Female				
	1–24	25–49	A	AA	Total	1–24	25–49	A	AA	Total
4–5	3.39	3.50	3.00	3.50	3.35	3.64	3.21	3.75	3.70	3.60
6–7	5.21	4.86	4.38	4.50	4.80	4.67	4.14	3.88	4.25	4.18
8–9	6.00	4.93	5.30	4.83	5.34	6.25	5.06	4.50	5.70	5.50

tends to be the case for the low-frequency words at all word lengths.

Number of different abbreviations. To assess variation in generation of the abbreviations, the number of different abbreviations was tabulated for the various conditions. Given the eight *S*s assigned to each treatment condition, eight was the maximum possible number of different abbreviations. Table 2 shows that the median number of abbreviations tends to decline with increasing T-L frequency. However, a reversal often occurs at the AA frequency, and in many instances the differences are trivially small. The T-L frequency was obviously not an important determinant of the number of different abbreviations. It can also be seen that the median number of different abbreviations increases regularly with word length and that, with some exceptions, female *S*s produced fewer different abbreviations than male *S*s. A chi-square analysis of word length and number of different abbreviations (3 × 8) demonstrated a significant relationship between the two variables: males, χ^2 (14) = 82.72, $p < .01$; females, χ^2 (14) = 68.47, $p < .01$. The T-L frequency was not analyzed because of the unsystematic nature of its effects.

EXPERIMENT II

This experiment sought to determine how well *S*s could explicitly use the abbreviation principles to construct word abbreviations. Although *S*s might demonstrate perfect consistency in the use of a given rule, it seemed more likely that, despite instructions to the contrary, they would employ whichever principle was best suited to the word at hand. More specifically, it was hypothesized that the effects of language properties, such as letter-sequence habits and redundancy, would tend to override the effects of the abbreviation principles whenever the principles produced an abbreviation different from one dictated by previous language experience.

Method

Subjects. The *S*s were 45 male and 45 female undergraduate students enrolled in the introductory psychology course at the University of Georgia. They participated to fulfill a course requirement.

Design and procedure. The present study consisted of three replications of Experiment I, each under different instructions. The three instructional conditions required independent *S*s (15 males and 15 females in each group) to construct abbreviations for the words of Experiment I making use of one of

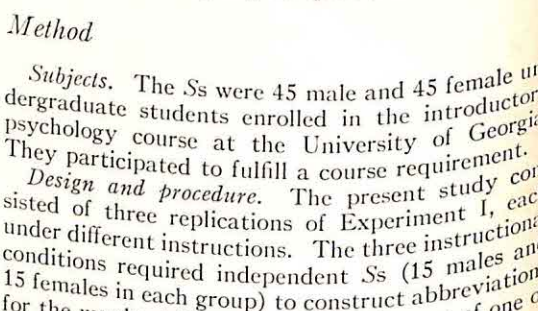


FIG. 2. Mean percentage of letters used in general abbreviations as a function of word length and Thorndike-Lorge frequency (Experiment II).

the three principles. The instructions defined and gave examples of the abbreviation principle *S*s were to employ. The *S*s given the joint principles were told that they could use truncation or contraction or both in devising their abbreviations. All other conditions and procedures were identical to those of Experiment I.

Results

Number of letters. The mean percentage of letters employed in the general abbreviations is shown in Figures 2 and 3 as a function of word length, T-L frequency, and instructional condition. It can be seen that the percentage of letters declines with increasing word length and T-L frequency. These effects are nearly identical to those of Experiment I. Again, the mean actual number of letters may be estimated from Figures 2 and 3 by multiplying the plotted values by 4.5, 6.5, or 8.5. An analysis of variance indicated that the effects of word length, $F\ (2, 72) = 7.40$, $p < .005$; T-L frequency, $F\ (3, 216) = 55.21$, $p < .001$; and their interaction, $F\ (6, 216) = 2.71$, $p < .025$, were statistically significant. The analysis also showed that the Word Length × T-L Frequency × Sex interaction was a significant source of variation, $F\ (6, 216) = 2.15$, $p < .05$. The remaining variability in the data was produced by the instructional conditions, $F\ (2, 72) = 10.69$, $p < .001$. Figure 3 shows that the group given contraction instructions used a larger percentage of the available letters than either of the other groups. The truncation and truncation–contraction *S*s employed nearly the same percentage of letters. With the 4–5 and 6–7 letter words, the contraction *S*s used approximately 5% more letters than the other groups; however, with 8–9 letter words the difference between the contraction *S*s and the other *S*s was close to 10%. Most of the larger difference is due to a decline in the percentage of letters used by the truncation and truncation–contraction *S*s.

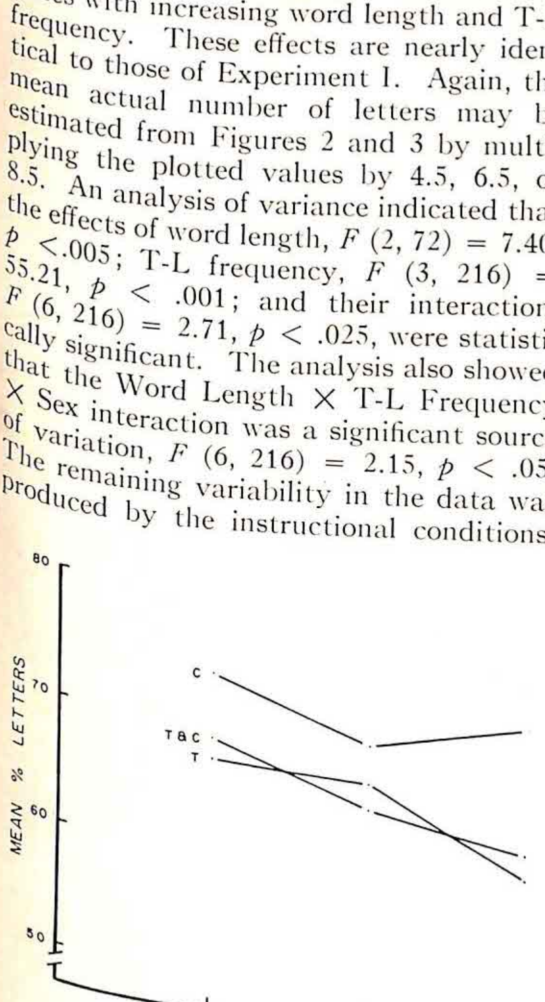


FIG. 3. Mean percentage of letters used in general abbreviations as a function of word length and instructional condition (Experiment II). (C = contraction; T = truncation; and T&C = truncation and contraction.)

Frequency of abbreviation principles. The frequency of use of the principles is shown in Table 3 as a function of word length, T-L frequency, instructional group, and sex for the general abbreviations. The results demonstrate that contraction once again was the preferred principle under most conditions of the experiment. With some exceptions, truncation was chosen over contraction only when *S*s were explicitly instructed to use truncation.

The first of several chi-square analyses evaluated the 3 × 3 tables for abbreviation principle and instructional condition of each sex (M or F) at each word length (4–5, 6–7, 8–9). The tests indicated that a significant association ($p < .01$) was present in each of the six contingency tables, $\chi^2\ (4) = 40.81$ (4–5 M), 42.55 (4–5 F), 61.35 (6–7 M), 57.53 (6–7 F), 72.05 (8–9 M), 97.28 (8–9 F). These results confirm the choice of contraction as the preferred principle. However, it must be kept in mind that truncation was used most frequently by the group instructed to construct truncation abbreviations. The overall superiority of contraction is mainly the result of its very heavy use in the contraction and truncation–contraction groups. Not surprisingly, the results of the truncation–contraction group parallel closely those of Experiment I. Given a free choice between truncation and contraction (the conditions of Experiment I), both males and females clearly preferred contraction for the shorter words, but chose truncation to abbreviate

TABLE 3

NUMBER OF WORDS ABBREVIATED AS A FUNCTION OF ABBREVIATION PRINCIPLE, WORD LENGTH, THORNDIKE-LORGE FREQUENCY, SEX, AND INSTRUCTIONAL CONDITION

Thorndike-Lorge frequency	Abbreviation principle	Instruction group							
		Male				Female			
		T	C	T + C	Total	T	C	T + C	Total
		Length 4–5							
1–24	T	9	3	1	13	9	2	4	15
	C	9	15	15	39	9	16	14	39
	TC	0	0	2	2	0	0	0	0
25–49	T	11	2	3	16	9	1	0	10
	C	7	16	14	37	9	17	18	44
	TC	0	0	1	1	0	0	0	0
A	T	13	2	3	18	7	1	2	10
	C	5	15	14	34	10	17	16	43
	TC	0	1	1	2	1	0	0	1
AA	T	9	4	8	21	8	0	3	11
	C	4	13	9	26	9	16	15	40
	TC	5	1	1	7	1	2	0	3
		Length 6–7							
1–24	T	14	4	4	22	7	1	2	10
	C	3	13	14	30	11	17	16	44
	TC	1	1	0	2	0	0	0	0
25–49	T	16	4	6	26	15	0	3	18
	C	2	12	11	25	3	14	11	28
	TC	0	2	1	3	0	4	4	8
A	T	17	3	7	27	12	2	10	24
	C	1	12	9	22	4	16	7	27
	TC	0	3	2	5	2	0	1	3
AA	T	12	5	6	23	10	0	3	13
	C	6	12	12	30	7	16	13	36
	TC	0	1	0	1	1	2	2	5
		Length 8–9							
1–24	T	6	0	4	10	13	0	1	14
	C	11	18	12	41	5	18	14	37
	TC	1	0	2	3	0	0	3	3
25–49	T	13	0	13	26	17	0	8	25
	C	5	18	5	28	1	18	10	29
	TC	0	0	0	0	0	0	0	0
A	T	14	1	13	28	15	2	13	30
	C	4	17	5	26	1	15	5	21
	TC	0	0	0	0	2	1	0	3
AA	T	12	1	13	26	17	3	13	33
	C	6	17	5	28	1	15	4	20
	TC	0	0	0	0	0	0	1	1

Note. T = truncation, C = contraction; and TC = truncation-contraction.

the longer words. However, it is also clear that within limits, *S*s can follow instructions to use particular abbreviation rules.

A second analysis evaluated the relationship between word length and abbreviation principle (T, C, or T + C) for each of the instructional conditions. Chi-square tests demonstrated a significant association in each of the 3 × 3 tables, χ^2 (4) = 18.92 (T), 20.78 (C), 51.14 (T + C), $p < .01$. Again, it was shown that contraction was preferred except under the conditions described above.

Chi-square tests of the 4 × 3 tables for

TABLE 4

MEDIAN NUMBER OF DIFFERENT GENERAL ABBREVIATIONS AS A FUNCTION OF WORD LENGTH, THORNDIKE-LORGE FREQUENCY, SEX, AND INSTRUCTIONAL CONDITION

Instructional condition and word length	Thorndike-Lorge frequency									
	Male					Female				
	1–24	25–49	A	AA	Total	1–24	25–49	A	AA	Total
Truncation (T)										
4–5	2.83	2.88	3.28	3.21	3.04	2.79	3.64	2.83	3.28	3.12
6–7	3.25	2.83	3.00	3.64	3.14	3.06	2.70	2.90	3.50	3.10
8–9	4.10	3.64	3.50	3.83	3.75	3.40	3.14	3.17	3.50	3.27
Contraction (C)										
4–5	2.62	2.30	2.10	2.38	2.32	3.00	3.10	3.21	3.50	3.19
6–7	4.06	4.04	3.88	3.88	3.97	3.50	3.83	3.61	3.28	3.59
8–9	3.96	4.21	3.60	4.28	3.99	3.96	4.17	4.00	3.90	4.02
T + C										
4–5	2.83	3.36	2.90	2.50	2.96	2.70	2.88	2.50	2.80	2.74
6–7	3.75	3.38	3.50	3.36	3.50	3.07	3.12	3.39	3.00	3.17
8–9	4.30	3.62	3.83	3.30	3.77	4.90	4.25	4.10	3.72	4.30

T-L frequency and abbreviation principle indicated a significant association in the truncation, χ^2 (6) = 22.62, $p < .01$, and truncation–contraction groups, χ^2 (6) = 27.46, $p < .01$, but not in the contraction condition, χ^2 (6) = 6.00, $p > .05$. Inspection of the individual tables suggested that the relative use of the three principles was about the same for all T-L frequencies in the contraction group. In both the truncation and truncation–contraction groups, the truncation rule was used more frequently and the contraction rule less frequently with increasing T-L frequency. These changes (and lack of change) seem to account for the results of the chi-square tests.

Finally, no tests were performed on the contingency table for sex and abbreviation principle because the same words were used with both groups, and thus violated the independence assumption of chi-square. However, the data showed that, contrary to Experiment I, females tended to use contraction somewhat more often than males. Both sexes constructed an equal number of abbreviations with the joint principles of truncation and contraction.

Number of different abbreviations. The median number of different general abbreviations is presented in Table 4 as a function of the experimental variables. The T-L frequency once again did not have a consistent effect on the abbreviations. In many cases, the number of abbreviations declined with increasing T-L frequency, but there are also contrary instances or an inconsistent trend. As in Experiment I, no analysis of this variable was performed. Word length, however, with one exception did produce regular increases in the median number of different abbreviations. Separate chi-square tests of the 3 × 5 tables for word length and number of abbreviations demonstrated a significant association between the two factors under all conditions except one, χ^2 (8) = 28.54 (M – T), 94.85 (M – C), 27.61 (F – C), 32.65 (M – T + C), 86.40 (F – T + C), $p < .01$. The exception was the abbreviations of female *S*s in the truncation group, χ^2 (8) = 7.97, $p > .05$.

Differences attributable to the instructional conditions were also highly consistent. Except in one case, the median number of abbreviations was higher in the contraction than in the truncation group. An opposite effect was found in the abbreviations of male *S*s given the 4–5 letter words. However, the results of the truncation–contraction group were mixed. The number of abbreviations was sometimes less, sometimes more, or sometimes intermediate to that of the other groups. Nevertheless, chi-square tests indicated a reliable relationship between instructional condi-

TABLE 5

MEAN CORRECT RECONSTRUCTIONS AS A FUNCTION OF WORD LENGTH, THORNDIKE-LORGE FREQUENCY, AND SEX

Word length	Thorndike-Lorge frequency							
	Male				Female			
	1–24	25–49	A	AA	1–24	25–49	A	AA
4–5	6.67	9.07	11.13	10.27	10.40	10.20	12.93	14.07
6–7	10.33	13.47	10.67	13.73	10.53	13.47	13.40	11.80
8–9	10.67	9.73	10.53	13.73	9.20	11.80	11.53	14.80

tion and number of different abbreviations, χ^2 (8) = 29.51 (4–5 M), 35.53 (6–7 M), 17.76 (6–7 F), 16.29 (8–9 M), 35.06 (8–9 F), $p < .05$, in all cases except the abbreviations constructed by female *Ss* given 4–5 letter words, χ^2 (8) = 14.72, $p > .05$.

EXPERIMENT III

The last study sought to test the adequacy or validity of the abbreviations constructed in the first experiment. Although an attempt was made in Experiment I to minimize idiosyncratic responses by requiring both personal and general abbreviations for each word, a more definitive test would be provided by asking other *Ss* to reconstitute the stimulus words from the abbreviations.

Method

Subjects. The *Ss* were 45 male and 45 female undergraduate students recruited mainly from introductory philosophy and sociology classes. They were paid $1.50 for participation in the experiment.

Design and procedure. The plan of the study closely followed Experiment I except that the stimuli were abbreviations rather than words. Independent *Ss* (15 males and 15 females) were randomly assigned to each of the three word-length conditions, while all *Ss* were given the four levels of T-L frequency.

The stimulus material consisted of the modal general abbreviations produced at each word length in Experiment I. Separate abbreviations were employed with males and females because the data of Experiment I suggested sex differences. When a modal abbreviation was not available (e.g., eight different abbreviations for a given word), one was chosen randomly. The abbreviations were assembled in booklets and arranged in the same random order as the words in Experiment I. The booklets were prefaced with a set of written instructions.

The *Ss*, run individually or in small groups, were asked to examine a set of abbreviations constructed by other students and to write down the word they believed each abbreviation represented. The *Ss* were told the number of letters in the words and that the words were common nouns. They were encouraged to guess if they were not sure or were uncertain about a word. All other procedures and conditions were the same as before.

Results

The dependent variable was the number of correct reconstructions produced by each *S* in the three word-length conditions. Each word length contained 72 abbreviations, divided equally among the four levels of T-L frequency (18/level). Thus, 18 was the maximum score of each *S* in each combination of word length and T-L frequency. Table 5 presents the mean correct reconstructions as a function of sex, T-L frequency, and word length. An analysis of variance indicated that significant differences were produced by word length (WL), F (2, 84) = 10.91, $p < .001$; sex (S), F (1, 84) = 17.99, $p < .001$; T-L frequency (TL), F (3, 252) = 67.16, $p < .001$; WL × S, F (2, 84) = 6.90, $p < .005$; TL × WL, F (6, 252) = 14.28, $p < .001$; and TL × WL × S, F (6, 252) = 9.99, $p < .001$. Only TL × S was not significant. With some exceptions, the females were superior to the males at all word lengths, but especially so with the 4–5 letter words at the lowest T-L frequency (10.40 vs. 6.67). Reconstruction was best for the medium-length words, and only slightly worse for the 4–5 and 8–9 letter words. Ignoring the males on the low-frequency 4–5 letter words, the mean percentage of correct reconstructions was approximately 67% at all word lengths.

Despite the large apparent variability in Table 5, perhaps the most striking finding was that correct reconstructions tend to increase fairly systematically with increasing T-L frequency for both males and females. Disregarding sex and word length, the overall T-L effect ranges from 53% correct reconstructions for the low-frequency words to 72% for the AA words, nearly a 20% increase. However, this increase is greater than 20% with the 8–9 letter words, smaller with the 6–7 letter words, and just about 20% for the short words. These differential effects of word length and T-L frequency are responsible for the significant TL × WL interaction, which is ordinal at 1–24 and 25–49, but disordinal for the A and AA words. The interaction becomes quite complicated when the sex factor is considered, thereby producing the significant triple interaction. The general superiority of the females had little influence on the effects of T-L frequency, but considerable influence on the word-length effects. In general, despite the results of the statistical analysis, the major finding was the systematic improvement of the reconstructions with increasing T-L frequency. The other variables and the various interactions acted mainly to produce variability in this finding.

General Discussion

In both Experiments I and II, the systematic decline in the percentage of letters with increasing word length apparently reflects the greater redundancy of longer words. With short words, omission of more than one or two letters would produce an abbreviation nearly meaningless to both the creator and other persons. With longer words, however, a larger percentage of the letters may be removed without vitiating the abbreviation. This interpretation of word length is also consistent with the finding that the actual number of letters used in the abbreviations increased only about one letter for each longer word-length condition (3–5 letters). These small increases in actual letters suggest that 3–5 letters may represent optimal abbreviation length. The fact that abbreviations of low-frequency words require a somewhat greater percentage of letters than high-frequency words suggests differential familiarity effects. The *S*s, being less familiar with low-frequency words and letter combinations inherent in these words, retained a larger percentage of the letters to protect the abbreviations from misinterpretation, and thus ensured that they would be meaningful to themselves and others. With A and AA words, these problems are less critical and a larger percentage of letters was omitted. On the other hand, despite the statistical differences produced by the frequency variable, the small absolute difference (about 5%) between the 1–24 and AA words is probably of little practical value.

Although the contraction principle was clearly preferred under the free-choice conditions of Experiment I, truncation was extremely important with the longer words. Truncation was apparently more popular with the longer words because the ends of the words were more predictable than the middle sections. Being more predictable, removing the end letters produced less degradation than deleting the middle items. However, the question may be asked why the ends were more predictable than the middle of the words. In an analysis of intraword constraint, Garner (1962) distinguishes between unilateral and bilateral constraint and states that the latter is the more important for the understanding of language. In the bilateral case, sequential constraint is greatest in the middle and least at the ends of a word. Constraint is maximal for middle letters because these letters are influenced by both forward and backward relations with other letters. In contrast, unilateral constraint, the constraint exercised from left to right in a word, grows (to a maximum of 50%) as a function of increasing word length. Furthermore, unilateral constraint can theoretically be only half as great as bilateral constraint. Given the importance of bilateral constraint and given the free-choice conditions of Experiment I, all abbreviations should have followed the contraction principle. And yet, the preferred use of truncation with the longer words implies that the constraint was unilateral rather than bilateral. In addition, *S*s in Experiment II, instructed to use a particular abbreviation rule, seemed able to overcome their own preferences (assumed comparable to those exhibited in Experiment I) and the effects of word length and T-L frequency. All these findings raise doubts about the applicability of the constraint model to word abbreviations. However, the model cannot be rejected because other variables may have influenced *S*s' choices. For example, the use of truncation with the longer words may be greatly dependent on the highly redundant suffixes commonly employed with long English words. Many of the 8–9 letter words did indeed possess such suffixes. In contrast, the 4–5 and 6–7 letter words had fewer such endings. Although further work would be necessary for a convincing demonstration, the suffixes are well organized and integrated as linguistic units, and thus are good candidates for omission in the construction of abbreviations. With shorter words, contraction is a sensible rule because it dictates the removal of highly constrained and familiar units such as vowels and bigrams. Because they are constrained and occur very frequently in everyday English, their omission has little effect on the meaningfulness of an abbreviation. Given the apparent difficulty of the constraint model with truncation abbreviations, it would be useful to determine whether the same problem occurs in the construction of abbreviations for other materials such as sentences. Does the model fail only with truncation abbreviations of words or does it have problems with other stimuli subject to abbreviation?

Variation in the number of different abbreviations was mainly a function of word length, with a greater variety of abbreviations produced by the longer words. Presumably the longer words offered more opportunities for the omission of noncritical letters than did shorter words. Often, for a given word, the differences among the abbreviations were the result of omitting a two-letter vs. a three-letter suffix or omitting a common suffix plus one of several interior letters. The possible ways of deleting letters are simply more numerous in long words, and it is not unreasonable to obtain a greater variety of abbreviations.

On an absolute basis, the reconstruction efforts in Experiment III were not spectacular, but, given the variability of the abbreviations, 67% correct reconstructions seems very good. Many of the abbreviations were based on only two or three *S*s, and thus not likely to engender high agreement among *S*s of Experiment III. An unexpected finding was the rather strong influence of T-L frequency on the reconstruction process. In Experiments I and II, the T-L effects were consistently ambiguous or had minimal influence on the abbreviations. On the other hand, in trying to restore missing letters in mutilated words (the task of Experiment III), the use of T-L frequency is a reasonable strategy, and probably the one with which *S*s have had most experience as language users. However, in the construction of abbreviations, T-L frequency is just not very relevant; it provides little guidance in the deletion of the letters. Word length is much more important because an abbreviation, by definition, requires a reduction in the total number of letters. A "good" abbreviation is one with the fewest number of letters and which is still meaningful to the user.

Not surprisingly, abbreviations for personal use require fewer letters than those designed for other persons, but the close parallel in the functions of the two types of abbreviation was not expected. The parallel results pose an interpretative problem which is made more difficult by a methodological error. On the one hand, the results suggest that similar processes govern both types of abbreviation. However, the parallel functions may also be a consequence of asking the same *S*s to produce both types of abbreviations at the same time. Another study is needed to evaluate the confounding which occurred in the present studies. The sex differences in all three studies are also difficult to interpret. They may genuinely reflect the commonly accepted view that females are superior to males on linguistic tasks or they may be artifacts of the experimental procedures. In the latter case, for example, the use of separate abbreviations for the two sexes in Experiment III is probably responsible in part for the observed differences and for the tremendous variability in the overall results.

These studies were begun in response to some practical questions concerning the best way to abbreviate everyday words. Although some tentative answers have been provided, the study of abbreviations may also have theoretical value. In addition to their relevance for Garner's (1962) constraint hypothesis, they may provide another approach to the investigation of coding processes. The construction of abbreviations can be viewed as a recoding or chunking process which serves to reduce code redundancy (Murdock, 1968) and thereby minimize the demand on central processing capacity. The fact that *S*s in the present studies used only 3–5 letters to construct abbreviations of words consisting of 4–9 letters is consistent with this view. Perhaps more importantly, the study of abbreviations might produce base-line information about what *S*s actually do when they encode a word for storage and retrieval in memory tasks. Although it is clear that *S*s can use a variety of stimulus attributes in performing coding operations (Underwood, 1969; Wickens, 1970), it may be sufficient in many situations to simply omit some of the elements, i.e., create an abbreviation. Furthermore, abbreviation-type codes have the virtue of being well practiced and understood by most users of written language. This suggestion does not deny *S*'s ability to use various attributes, but only states that the typical *S* may choose to employ fairly simple and obvious coding procedures when given a free choice. And lastly, it can be noted that, despite the imprecise results of Experiment III, *S*-generated abbreviations probably have considerable potential as retrieval cues.

REFERENCES

Barrett, J. A., & Grems, M. Abbreviating words systematically. *Communications of the Association of Computing Machinery*, 1960, **3**, 323–324.

Bourne, C. P., & Ford, D. F. A study of methods for systematically abbreviating English words and names. *Journal of the Association for Computing Machinery*, 1961, **8**, 538–552.

Garner, W. R. *Uncertainty and structure as psychological concepts.* New York: Wiley, 1962.

Luhn, H. P. *Superimposed coding with the aid of randomizing squares for use in mechanical information searching systems.* Poughkeepsie, N. Y.: IBM Product Development Laboratory, IBM Corporation, 1956.

Murdock, B. B., Jr. Decoding as a function of the number of bits per chunk. *Journal of Experimental Psychology*, 1968, **78**, 1–7.

Thorndike, E. L., & Lorge, I. *The teacher's word-book of 30,000 words.* New York: Columbia University, Teachers College, Bureau of Publications, 1944.

Underwood, B. J. Attributes of memory. *Psychological Review*, 1969, **76**, 559–573.

Webster's New World Dictionary of American Language. (College ed.) Cleveland: World Publishing, 1966.

Wickens, D. D. Encoding categories of words: An empirical approach to meaning. *Psychological Review*, 1970, **77**, 1–15.

(Received July 17, 1972)

Journal of Experimental Psychology
1973, Vol. 98, No. 2, 362–367

PRESENTATION MODALITY, DISTRACTOR MODALITY, AND PROACTIVE INTERFERENCE IN SHORT–TERM MEMORY[1]

RONALD H. HOPKINS,[2] RICHARD E. EDWARDS, AND CHERI L. COOK

Washington State University

Two experiments were conducted to investigate the previously reported asymmetrical release from proactive interference following a shift in mode of presentation of the memory items. The asymmetry was not changed by warning *S* of the mode of presentation at the beginning of each trial, but the direction of the asymmetry was reversed by changing the mode of presentation of the interpolated task. In addition, it was found that overall recall is better when the modes of presentation of the distractor and to-be-remembered items are different than when those modes are the same, and that changes in recall are not simply due to compensatory changes in filler-task performance. It was concluded that auditorily and visually presented verbal materials can be encoded differently and that the locus of these differences is in the short-term store.

The encoding of verbal materials has been intensively studied in the past few years. One topic of particular interest is encoding in short-term memory, that is, for retention over intervals of a few seconds. The importance of auditory and/or articulatory factors has been well established (Conrad, 1964), and a number of theorists have emphasized these factors in their descriptions of short-term memory (Atkinson & Shiffrin, 1968; Sperling, 1967). However, there is accumulating evidence that the short-term memory code depends on the mode of presentation (auditory, A, or visual, V) of the memory items, and the code sometimes preserves much of the visual information of a V stimulus (Kroll, Parks, Parkinson, Bieber, & Johnson, 1970; Murdock & Walker, 1969).

Recently, Hopkins, Edwards, and Gavelek (1971) used the Wickens (1970) release-from-proactive-interference (PI) paradigm to test for differential encoding of A and V materials. In its simplest form, the assumption is that if A and V materials are encoded differently they should not interfere with one another; thus, a shift from one mode to the other should produce a recovery from any accumulated PI. Hopkins et al. found that a shift of A presentation following three V trials produced an abrupt improvement in performance, consistent with the hypothesis of modality-specific encoding; however, a comparable shift in the A to V direction did not produce the release from PI. As was pointed out in that article, encoding differences could hardly be expected to produce a release from PI in one direction, but not in the other. Consequently, if the release from PI is to be attributed to differential encoding of A and V materials, another factor must be involved. The experiments reported here were designed to determine the variables producing the asymmetrical release from PI following a mode shift.

EXPERIMENT I

Hopkins et al. (1971) suggested the "attention hypothesis" as one possible explanation for the asymmetrical release from PI in their experiments. According to this hypothesis, *S*s in the A to V shift condition stopped attending to the V display after several successive A trials, even though warned initially that a shift in modality might occur. If this hypothesis were correct, many *S*s would not have perceived the V materials of the A to V shift and, consequently, could not have demonstrated the release from PI. Although internal analyses of the earlier data did not support the attention hypothesis, it was not tested directly. In this experiment, an attempt was made to insure that *S* attended to the visual display. He was simply informed at the beginning of each trial of the mode of presentation to be used.

---

[1] This research was supported by Grant MH-20232-01 from the U.S. Public Health Service, by Research Grant Award OEG-X-71-0032(508) from the U.S. Office of Education, and by Washington State University Graduate School Research Funds. The second author was a U.S. National Aeronautics and Space Administration Fellow during the conduct of this research.

[2] Requests for reprints should be sent to Ronald H. Hopkins, Department of Psychology, Washington State University, Pullman, Washington 99163.

Method. Each *S* received four successive trials in a distractor-type short-term retention task. The conditions of study remained constant for each *S* over the first three trials. The fourth trial was a shift trial to test for release from accumulated PI. Four independent groups were formed by the factorial combination of mode shift (present or absent) and mode of presentation on the fourth trial (A or V). A pair of letters will be used to designate each of these conditions: The first letter (A or V) refers to the mode of presentation on the first three trials, the second letter (A or V) to the mode of presentation on the fourth trial.

The to-be-recalled study items were triads of high-frequency animal names from the Battig and Montague (1969) norms. Two different lists were prepared by independently forming four triads from the word pool; these lists were counterbalanced across conditions. The members of a triad were not related in any obvious way (other than category membership), and the mean normative frequency was approximately the same for all triads. The distractor or filler materials on each trial were a randomly selected series of six signed digits from the pool +1, −1, +2, and −2. When a +1 or a −2 was presented, *S* was to say "A," and when a −1 or a +2 was presented, *S* was to say "B."

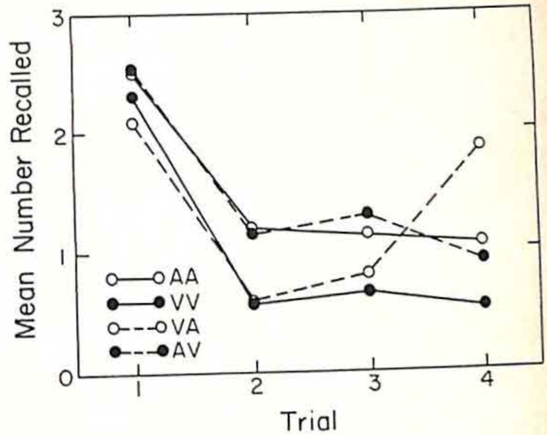


FIG. 1. Mean recall performance on Trials 1 to 4 of Experiment I. (A = auditory; V = visual; first letter = mode of presentation on Trials 1 to 3; second letter = mode of presentation on Trial 4.)

All V materials were presented on a rear-projection screen by alternating Kodak Carousel projectors. Auxiliary knife shutters were programmed so that all V exposures were of .5 sec. duration, approximately the duration of the A stimuli. The A materials were presented by tape recorder via headphones. The A and V presentations were sychronized by a digital programmer with the program stored on the extra channel of the stereo tape deck. Whenever an A item was presented, a blank slide was flashed on the screen.

Each trial begain with a warning signal which informed *S* of the mode of presentation for the upcoming trial. For V trials this signal was the word "visual" presented simultaneously on the screen and over the headphones; for A trials the word "auditory" was presented simultaneously in both modalities. After 1.5 sec., the study words were presented one at a time at intervals of .75 sec., onset to onset. The first filler item was presented .75 sec. after the onset of the last study word. The six filler items were then presented at intervals of 2.50 sec., onset to onset. As in the Hopkins et al. (1971) experiments, the filler items were presented in the V modality on all trials. After the last filler item, the word "recall" was presented on the screen and *S* had 2.50 sec. to say aloud the three study words. There was no intertrial interval; the warning signal for the next trial was presented immediately after the recall period. This procedure was repeated without interruption for four trials.

The *S* was informed that once a triad had been tested it would not occur again. He was also instructed to recall the words in the same order as they had been presented, and that he could guess if uncertain.

The *Ss* were 128 Washington State University undergraduates who participated in partial fulfillment of the course requirements for introductory psychology. Thirty-two *Ss* were assigned to each of the four groups according to a block-randomized ordering.

Results and discussion. Errors in order of recall occurred on only 6.4% of the occasions where there was possibility for order error, so only item recall will be considered here. The mean numbers of words correctly recalled are presented in Figure 1 as a function of trials. Performance on Trials 1 to 3 was analyzed first. As may be seen in the figure, recall declined markedly from the first to the third trial, F (2, 248) = 104.05, $p < .001$, indicating rapid development of substantial PI. The only other significant source of variance was that A items were recalled better than V items, F (1, 124) = 22.65, $p < .001$. This difference is consistent with our previous results (Hopkins et al., 1971, Experiments I and II), as well as with the differences generally obtained by other investigators (e.g., Grant & McCormack, 1969).

Analysis of the shift results from Trial 4 would ordinarily involve the comparison of Condition VA to its control, Condition AA, and the comparison of Condition AV to its control, Condition VV. However, this analysis could be misleading because the apparent Trial 4 superiority of Condition AV to Condition VV may be entirely due to carry-over from the main effect of mode on Trials 1 to 3. Consequently, absolute release from PI was evaluated by using recovery scores. The recovery score was computed for each *S* by subtracting Trial 3 recall from Trial 4 recall. The mean recovery scores may be estimated from Figure 1 and are 1.03, $-.06$, $-.41$, and $-.13$ for the VA, AA, AV, and VV conditions, respectively. The only one of these mean recovery scores which was significantly different from 0.0 was that for Condition VA, $t\ (31) = 4.30$, $p < .001$. It is concluded that under these conditions there is a release from PI for a VA shift but not for an AV shift. Therefore, the warning signal was unsuccessful in producing a symmetrical release from PI.

Distractor-task performance was also analyzed. Murdock (1966) has suggested that decrements in short-term recall, such as the development of PI, are often accompanied by gains in performance on the fillter task. Although more recent evidence (Crowder, 1968; Frost & Jahnke, 1968) indicates that changes in recall are not entirely due to compensatory changes in performance on the interpolated task, the possible influence of this factor was examined. When pooled over the four shift conditions, the mean numbers of correct responses on the filler task were 3.62, 4.41, 4.44, and 4.53 for Trials 1 to 4, respectively. The performance increase over trials was significant, $F\ (3, 372) = 24.58$, $p < .001$. However, this improvement seems to have been due to general practice effects rather than to the experimental manipulations. This is evidenced by the fact that neither shift nor Trial 4 mode was a significant source of variation, either in main effect or in interaction with trials. Thus, the development of PI might be attributed to a corresponding improvement in classification performance, but the modality effect and the release from PI cannot have been due simply to compensatory changes in filler-task performance.

EXPERIMENT II

Since all of our previous experiments had used an interpolated task presented in the V mode, this experiment was simply designed to assess the role of filler modality in this situation. There are a number of indications in the literature that a distractor task interferes more with retention when it is presented in the same mode as the study items than when the study and distractor modes differ (Kroll et al., 1970; Margrain, 1967; Parkinson, 1972). Thus, it was predicted that the direction of the main effect of modality would depend upon the filler modality; furthermore, it seemed likely that the filler modality might influence the asymmetry in the release from PI.

Method. Eight independent groups were formed by the factorial combination of mode shift (present or absent), mode of the memory items on the shift trial (A or V), and mode of the filler task (A or V). This experiment was conducted simultaneously with Experiment I, with only two differences in materials and procedure. First, the warning signal at the beginning of each trial was replaced by the word "study" for continuity with the procedure of Hopkins et al. (1971). Second, the filler items were presented in the assigned modality via the screen or headphones. The *S* always verbalized the response to each filler item. The 256 *S*s were selected from the same pool as for Experiment I, and 32 were randomly assigned to each of the eight groups.

Results and discussion. As in Experiment I, order errors occurred on only 6.9% of the opportunities and will not be considered further. The mean recall scores are presented in Figure 2, with the curves for the V filler in the left panel and for the A filler in the right panel. Considering first the data for the development of PI, performance on the first three trials was analyzed with filler mode, presentation mode, shift, and trials as variables. Obviously, recall declined across the first three trials, $F\ (2, 496) = 157.48$, $p < .001$. Overall, recall was better with the V filler than with the A filler, $F\ (1, 248) = 16.94$, $p < .001$, and recall was better for words presented

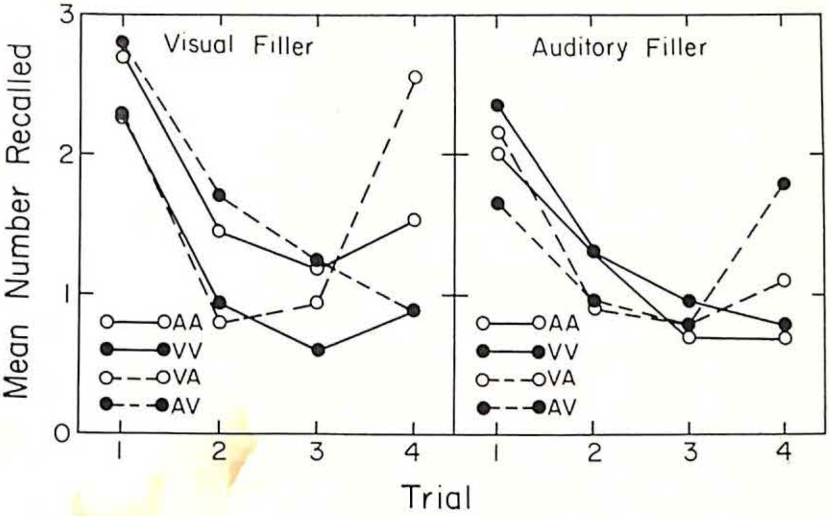


FIG. 2. Mean recall performance on Trials 1 to 4 of Experiment II. (A = auditory; V = visual; first letter = mode of presentation on Trials 1 to 3; second letter = mode of presentation on Trial 4.)

in the A mode than in the V mode, F (1, 248) = 8.71, $p < .005$. However, presentation mode and filler mode interacted reliably, F (1, 248) = 33.40, $p < .001$, such that with a V filler, recall of A items was better than that of V items, F (1, 248) = 38.12, $p < .001$, but with an A filler, recall of V items was better than that of A items, F (1, 248) = 4.00, $p < .05$. This interaction supports the findings of other investigators that a distractor interferes more with retention when it is presented in the same mode as the study item. Unfortunately, there was also an interaction between filler mode and shift, F (1, 248) = 6.86, $p < .01$. The nature of this interaction was such that the recall of shift groups ($\bar{X} = 1.21$) was poorer than that for no-shift groups ($\bar{X} = 1.44$) for the A-filler condition, but that the reverse was true for the V-filler condition ($\bar{X} = 1.63$ for the shift groups and $\bar{X} = 1.53$ for the no-shift groups). Since the shift and no-shift conditions were treated identically on the first three trials, the differences must be due to sampling error. Furthermore, although this interaction may have attenuated the release from PI in the A-filler conditions and enhanced it in the V-filler conditions, it almost certainly could not account for the particular pattern of results obtained on Trial 4.

Turning to the shift data, the recall scores from Trial 4 were analyzed with presentation mode, shift, and filler mode as variables. The three-way interaction of these variables was significant, F (1, 248) = 11.94, $p < .001$, so the data were analyzed separately for the V-filler and for the A-filler conditions. With a V filler, A presentation of the study words produced better recall than V presentation, F (1, 124) = 54.04, $p < .001$, and a mode shift produced better recall than no shift, F (1, 124) = 10.46, $p < .005$. However, as shown in Figure 2, the shift effect was confined entirely to the VA condition, resulting in a Mode × Shift interaction, F (1, 124) = 10.46, $p < .005$. For the V filler then, the results confirm those obtained in Experiment I and in our earlier experiments (Hopkins et al., 1971). With an A filler, V presentation of the study words produced better recall than A presentation, F (1, 124) = 5.11, $p < .05$, and a shift produced better recall than no shift, F (1, 124) = 16.55, $p < .001$. In this case, the Mode × Shift interaction was not reliable, F (1, 124) = 2.95, $p \simeq .09$. However, it is apparent in Figure 2 that recall was better in Condition AV than in the VV control, F (1, 124) = 16.74, $p < .001$, whereas recall in Condition VA was little, if any, better than that in the AA control,

$F\ (1, 124) = 2.76$, $p > .10$ These results were confirmed by analyses of the Trial 4 minus Trial 3 recovery scores. For the V filler, the mean recovery score was significantly different from 0.0 only on Condition VA, $t\ (31) = 8.36$, $p < .001$; for the A filler, the mean recovery score was significantly different from 0.0 only in Condition AV, $t\ (31) = 5.56$, $p < .001$. It is concluded that an asymmetrical release from PI is obtained with both an A filler and a V filler, and that the direction of the asymmetry can be manipulated by changing the mode of the filler task.

When pooled over shift conditions, the mean numbers of correct responses to the A filler items were 4.02, 4.12, 4.26, and 4.28 for Trials 1 to 4, respectively. The corresponding means for the V filler were 3.78, 4.26, 4.46, and 4.53, respectively. Filler performance improved significantly as a function of trials, $F\ (3, 744) = 12.25$, $p < .001$, and the rate of improvement was greater for the V filler than for the A filler, $F\ (3, 744) = 2.79$, $p < .05$. Although filler performance was slightly better for conditions receiving V presentation on Trial 4 than for those with A presentation, $F\ (1, 248) = 6.58$, $p < .025$, none of the other main effects or interactions was reliable. Thus, the release from PI cannot be due to a simple compensatory relationship between retention and filler performance; the improvement in filler performance probably resulted from a general practice effect which was somewhat more pronounced for the V-filler task than for the A-filler task.

GENERAL DISCUSSION

The results of these experiments are reasonably clear: (*a*) The earlier finding of a release from PI following a VA shift, but not an AV shift, can be replicated. (*b*) The asymmetry is not changed by warning *S* of the mode of presentation at the beginning of each trial; that is, the attention hypothesis is incorrect. (*c*) The direction of asymmetry in the release from PI can be completely reversed by presenting the interpolated task in the A mode rather than in the V mode. (*d*) When the distractor task is presented in the A mode, it interferes more with the retention of A memory items than with V items; presenting the distractor task in the V mode results in more interference with V memory items than with A items. (*e*) Changes in recall cannot be attributed simply to compensatory changes in performance on the filler task.

It is assumed that items which do not interfere with one another must, to some extent, have different memorial representations. Given this assumption, the present results suggest that A and V materials are, at least under some conditions, encoded differently. That is, there are two different indications that intermodal interference is lower than intramodal interference. First, recall is better when the modes of presentation of the distracter and to-be-remembered items are different than when those modes are the same. Second, under appropriate conditions, a release from PI is obtained following a shift in the mode of presentation of the memory items.

An important question, of course, is the locus of these effects. The theoretical implications would seem to be most interesting if the processes producing the observed effects are in the short-term store. It is reasonable to eliminate a precategorical store (Crowder & Morton, 1969; Neisser, 1967) as the possible locus on the grounds that the retention intervals used here exceed the lifetimes ordinarily associated with items in these precategorical stores. On the other hand, it has been suggested that some portion of the release from interference observed in this paradigm reflects long-term as opposed to short-term processes (Hopkins, Edwards, & Cook, 1972). In fact, Craik and Birtwistle (1971), using a free-recall procedure, concluded that the release from PI is due entirely to long-term processes. Nevertheless, most memory theorists assign rehearsal processes to the short-term store, and the results of Experiment II clearly show that a variable which affects rehearsal—the nature of the distractor task—has a profound effect on the release from PI. Thus, in contrast to the free-recall procedure, the release from PI obtained in the short-term distractor task and any inferred encoding differences primarily seem to reflect processes operating in the short-term store.

These experiments, then, are taken as evidence that A and V materials can have different representations in short-term memory. This is not to say that an auditory and/or articulatory code does not play a major role in verbal short-term memory, or even that the code for V items is visual or imaginal in nature. It is simply suggested that not all verbal items are stored as though they had been presented in the A mode, and that theories of memory will need to take account of this fact.

REFERENCES

Atkinson, R. C., & Shiffrin, R. M. Human memory: A proposed system and its control processes. In K. W. Spence & J. T. Spence (Eds.), *The psychology of learning and motivation: Advances in research and theory.* Vol. II. New York: Academic Press, 1968.

Battig, W. F., & Montague, W. E. Category norms for verbal items in 56 categories: A replication and extension of the Connecticut category norms. *Journal of Experimental Psychology Monograph*, 1969, **80**(3, Pt. 2).

Conrad, R. Acoustic confusion in immediate memory. *Quarterly Journal of Experimental Psychology*, 1964, **55**, 75–84.

Craik, F. I. M., & Birtwistle, J. Proactive inhibition in free recall. *Journal of Experimental Psychology*, 1971, **91**, 120–123.

Crowder, R. G. The relation between interpolated-task performance and proactive inhibition in short-term retention. *Journal of Verbal Learning and Verbal Behavior*, 1968, **7**, 577–583.

Crowder, R. G., & Morton, J. Precategorical acoustic storage (PAS). *Perception & Psychophysics*, 1969, **5**, 365–373.

Frost, R. F., & Jahnke, J. C. Proactive effects in short-term memory. *Journal of Verbal Learning and Verbal Behavior*, 1968, **7**, 785–789.

Grant, K. W., & McCormick, P. D. Auditory and visual short-term memory with successive syllable presentation in both modalities. *Psychonomic Science*, 1969, **17**, 341–342.

Hopkins, R. H., Edwards, R. E., & Cook, C. L. The dissipation and release of proactive interference in a short-term memory task. *Psychonomic Science*, 1972, **27**, 65–67.

Hopkins, R. H., Edwards, R. E., & Gavelek, J. R. Presentation modality as an encoding variable in short-term memory. *Journal of Experimental Psychology*, 1971, **90**, 319–325.

Kroll, N. E. A., Parks, T., Parkinson, S. R., Bieber, S. L., & Johnson, A. L. Short-term memory while shadowing: Recall of visually and aurally presented letters. *Journal of Experimental Psychology*, 1970, **85**, 220–224.

Margrain, S. A. Short-term memory as a function of input modality. *Quarterly Journal of Experimental Psychology*, 1967, **19**, 109–114.

Murdock, B. B., Jr. Measurement of retention of interpolated activity in short-term memory. *Journal of Verbal Learning and Verbal Behavior*, 1966, **5**, 469–472.

Murdock, B. B., Jr., & Walker, K. D. Modality effects in free recall. *Journal of Verbal Learning and Verbal Behavior*, 1969, **8**, 665–676.

Neisser, U. *Cognitive psychology.* New York: Appleton-Century-Crofts, 1967.

Parkinson, S. R. Short-term memory while shadowing: Multiple-item recall of visually and of aurally presented letters. *Journal of Experimental Psychology*, 1972, **92**, 256–265.

Sperling, G. Successive approximations to a model for short-term memory. *Acta Psychologica*, 1967, **27**, 285–292.

Wickens, D. D. Encoding categories of words: An empirical approach to meaning. *Psychological Review*, 1970, **77**, 1–15.

(Received July 17, 1972)

Journal of Experimental Psychology
1973, Vol. 98, No. 2, 368–374

EFFECTS OF BLANK-TRIAL PROBES ON CONCEPT-IDENTIFICATION PROBLEMS WITH REDUNDANT RELEVANT CUE SOLUTIONS[1]

MOSHE J. LEVISON[2] AND FRANK RESTLE

Indiana University

Levine's *blank-trial assumption* says that *S*s in a simple concept-identification experiment learn only on feedback trials, not on blank (nonfeedback) trials. If so, then blank trials could be a useful tool to track the course of learning in more complex concept problems. This experiment tests the extended blank-trial assumption by giving *S*s problems that can be solved by learning either or both of two relevant redundant cues (RRC). Four groups of college students were given the same series of feedback trials, double alternated with different sets of blank trials. It was found that *S*s made more errors in learning when cues were missing on blank trials. When the relevant cues were put in conflict on blank trials, or when only one of the two cues was shown, relatively few *S*s learned both cues. Thus, the blank trials are not a valid test of learning RRC solutions, since they may affect both the speed of learning and what is learned. The relationship between the present results and previous findings is discussed.

The blank-trial procedure, introduced by Levine (1966), has become the most popular paradigm to study the process of hypothesis (H) testing on a trial-by-trial basis during the acquisition phase of concept-identification (CI) problems. Levine presents structured sequences of training stimuli interspersed among the standard feedback trials; *S* responds to these, but feedback is withheld. Levine made two assumptions that permit him to infer *S*'s H from the response pattern to the blank-trial sequence. The *blank-trial assumption* states that when *S* receives no feedback, he maintains the same H for the next trial. Coding and processing of the stimulus occur only after feedback is given (Levine, 1966, p. 337). A second assumption that has since been revised was the *basic assumption* that a single H determines the responses to each sequence of blank trials. In the revised formulation (Levine, 1970), *S* may process several Hs at a time on the feedback trials; either a single "working H" or "majority rule" determines the overt response. Regardless of which response rule is adopted, the H set being tracked is never modified on blank trials. The stability of the H set following a blank-trial sequence distinguishes positive-feedback effects from blank-trial effects.

The blank-trial probe procedure has been extended to various learning situations other than discrimination learning and CI, including paired-associate learning (Izawa, 1967), auditory discrimination (Moore & Halpern, 1967), and probability learning (Levine, Leitenberg, & Richter, 1964; Richter & Levine, 1965). Another variation upon Levine's procedure in CI and other tasks has been to introduce a contingency between *S*'s responses and the blank trials to compare the effectiveness of positive, negative, and blank feedbacks (e.g., Bourne, Guy, & Wadsworth, 1967; Buchwald, 1959; Buss & Buss, 1956; Rimm, Roesch, Perry, & Peebles, 1971; Spence, 1970, 1972).

Except for the results of Izawa (1967), where test trials retarded forgetting, blank trials were found to influence learning rates primarily in cases where blank feedback was presented only if *S*'s response was incorrect (right vs. blank). When blank feedback followed only correct responses (blank vs. wrong) or followed either kind of response, learning rates were the same as for groups receiving no blank trials (right vs. wrong). These data are all consistent with the hypothesis that *S*s do not switch or add Hs on blank trials. If blank trials are presented following incorrect responses and *S* treats blank trials as positive feedback (not resampling Hs following either blank or positive feedback; cf. Spence, 1970), learning will be retarded until the tendency is overcome (Buchwald, 1959). The *S* must first learn to switch Hs after a blank trial before he can reach criterion. Unless (or until) this occurs, blank trials will affect learning. When veridical feedback follows all (or some) errors, blank trials do not prevent *S* from resampling Hs after errors and so do not affect learning.

---

[1] This work was supported in part by Grant MH-16817-03 from the U. S. Public Health Service. The authors thank Virginia Wasser for her assistance in running the experiment. Portions of these data were reported at the meeting of the Midwestern Psychological Association, Cleveland, May 1972.

[2] Requests for reprints should be sent to Moshe Levison, Department of Psychology, Indiana University, Bloomington, Indiana 47401.

While the CI and discrimination learning research using noncontingent blank-trial probes has provided support for Levine's assumptions, the problem solutions have all (with one single exception; i.e., Slaymaker, 1972) been rules involving only a single, simple H. In these problems (Berger, 1970, excepted) the blank-trial stimuli have come from the population of feedback-trial stimuli, and all the information about the problem is transmitted via the feedback. A strong test of the blank-trial assumption would be to test the effects of blank trials in a more complex CI problem where the stimuli displayed on blank trials would differ from those presented on the feedback trials. One concept-identification problem that would allow such a test for the blank-trial assumption is one having a relevant redundant cue (RRC) solution.

In a two-RRC problem the relevant dimensions are covaried on all trials and *S* can reach criterion in acquisition by attending to *either* or *both* of the redundant cues. The introduction of blank trials would allow *E* to track the solution process during acquisition and determine on each trial whether *S* has one or both relevant cues rather than relying upon postacquisition tests or transfer of training tests (Trabasso & Bower, 1968). The blank-trial procedure can measure *S*'s H set only if *S*s do not modify their H set as a result of the blank-trial stimuli. The reason for this restriction is that to obtain information as to which redundant cues are being used, the blank-trial stimuli have to dissociate the relevant cues, and the blank-trial stimuli would differ from the feedback-trial stimuli and would not follow the RRC rule. These differences may provide new information to *S* even though feedback is omitted. The blank-trial assumption, if true, predicts that the composition of the blank-trial stimuli should not affect the solution of the problem since *S* does not modify his H set on the blank trials. Any differences in the distribution of solution types for a two-RRC problem as a function of the nature of the blank-trial stimuli would invalidate the extension of the blank-trial procedure to RRC problems. The purpose of the present experiment was to train groups of *S*s on a two-RRC problem and intersperse blank-trial stimuli of different compositions to determine their effects, if any, on the solution of the problem.

METHOD

Subjects. A total of 494 undergraduate students enrolled in introductory psychology courses at Indiana University participated in the experiment as part of their course requirements.

Apparatus and materials. Stimuli were drawn on transparencies and projected on a screen by an overhead projector. Each stimulus was one combination of four binary dimensions: (*a*) shape (triangle or circle); (*b*) position of a dot (upper or lower half of the stimulus); (*c*) center sign (plus or minus); and (*d*) border (solid or dashed lines).

Procedure. The *S*s were run simultaneously in groups of 5–75 in a lecture hall seating 100. Responses were coded by *S*s on optically scored answer sheets. Standard instructions for redundant relevant cue conditions were read (Trabasso & Bower, 1968, pp. 72–73), and instructions for the blank-trial probes were like Levine's (1966, p. 333). In brief, *S*s were told they could learn a rule that would enable them to solve the problem and correctly classify each stimulus into one of two categories, alpha or beta. They were cautioned that feedback would not be given on each trial, but that they should respond on every trial even if they were only guessing. A sample stimulus was projected and *E* pointed out the two values of each of the four dimensions. The *S*s were told that the correct rule involved only the stimulus dimensions just described to them. The order of stimuli was randomized for each group of *S*s.

In every condition, *S*s received 96 trials, 48 reinforced trials and 48 blank trials in double alterna-

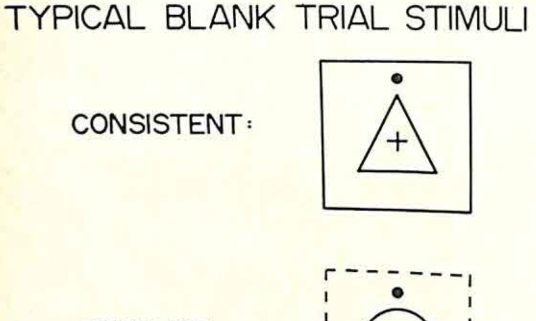


FIG. 1. Examples of the stimulus configurations shown on feedback and nonfeedback trials and the test-phase trials. (The redundant relevant cue [RRC] response rule was: dot above triangle = alpha; dot below circle = beta. The border and center sign were irrelevant dimensions. The consistent stimulus shows a dot above the triangle. Similar stimuli were shown on all feedback trials for all conditions. The conflict stimulus shows a dot above the circle and is not in accord with the RRC rule. The single-cue stimulus shows a triangle in isolation. The omit stimulus shows the dot below and the irrelevant dimensions, but the circle is deleted.)

tion. Stimuli were projected for 10 sec. followed by a 5-sec. intertrial interval when *E* gave the correct response on feedback trials. The present procedure departed from Levine's (cf. 1969) in several respects. Levine's *S*s were trained on simultaneous discrimination problems rather than on a successive problem. Levine tested each *S* individually on 16 problems, all having the same type of solution, after giving extensive pretraining and instructions. Levine carefully ordered the presentation sequence of both the feedback-trial stimuli and the blank-trial sequences with four blank trials following every feedback trial.

By using relatively fewer blank trials, we should reduce their total impact. The present experiment could not use *S*s who have been indoctrinated in a particular solution type—nor can any RRC experiment. There is no apparent reason for the blank-trial assumption to apply to simultaneous but not successive CI problems.

The four conditions differed only with respect to the particular stimuli shown during blank trials (see Figure 1). These were as follows:

1. Consistent: Blank-trial stimuli were identical to those presented on reinforced trials. This was the control condition.

2. Omit: The blank-trial stimuli each contained a value of one of the relevant dimensions, but the second relevant dimension was omitted. The particular relevant dimension presented varied from trial to trial. Both irrelevant dimensions appeared on every trial.

3. Conflict: The blank-trial stimuli contained values of both relevant dimensions, but the values were not redundant. The relevant dimensions were allowed to conflict and were not in accord with the rule applying on feedback trials. Irrelevant dimensions were present.

4. Single-cue: The blank-trial stimuli consisted of a single value of one of the dimensions, either relevant or irrelevant.

Following the 96 trials of acquisition, all *S*s were given 48 nonfeedback test trials to determine which relevant cue or cues they had learned in acquisition. The test-phase stimuli for all conditions were the same as those presented on the blank trials under the omit condition described above.

Following the test phase, *S*s were administered a questionnaire asking them how they had solved the problem and how the blank-trial stimuli had influenced their behavior.

Dependent variables were: (*a*) the number of errors and trial of last error before a criterion of 10 consecutive correct reinforced responses in the acquisition phase of the experiment (blank trials were excluded); (*b*) number of redundant relevant cues (one or two) learned by *S*, as inferred from the test-phase responses; and (*c*) responses to the questionnaire, as to whether *S*s used one or two cues, whether they believed blank trials were relevant or irrelevant to the problem, whether they believed their blank-trial responses were correct or incorrect, and whether they changed their H set during the blank trials by adding, deleting, or switching cues.

RESULTS

Learning. The four conditions tested did not differ in the number of *S*s who reached criterion in acquisition, χ^2 (3) = 3.53, or in the mean trial of last error, *F* (3, 490) = .8. The four conditions' mean total errors were significantly different among themselves, *F* (3, 490) = 2.89, $p < .035$ (see Table 1).

On all three measures of acquisition, the *S*s in the omit and single-cue conditions had slightly greater difficulty in solving the problem than the *S*s in the consistent and conflict conditions who were virtually identical in acquisition performance. Removing parts of stimulus configurations on blank trials had the general effect of slightly retarding learning, increasing the total errors by one or two.

Test trials. The major finding lay not in how quickly *S*s solved the problem, but in the test-phase results indicating *how* *S*s in each condition solved the problem. It was here that the differences among the experimental conditions emerged. In the consistent condition, which serves as the base line, .586 of *S*s who solved the problem learned both redundant relevant cue solutions to the problem. This was exceeded, though not significantly, χ^2 (1) = 1.25, by the single-cue condition: .680 of *S*s in this condition learned both cues. In the omit condition, .404 used both relevant cues, and in the conflict condition, .299 solved the problem by using both redundant cues. These latter groups did not differ between themselves, χ^2 (1) = 2.97, $p < .10$, but they differed, χ^2 (1) = 24.65, $p < .001$, from the consistent and the single-cue conditions (see Table 2).

A total of 76 of the 494 *S*s responded at chance levels to both of the relevant dimensions in the test phase, and their data were not included in the above analyses. There were no differences among conditions in the numbers of these nonsolving *S*s, χ^2 (3) = 2.53.

Postexperimental questionnaire. All four groups of *S*s responded the same overall to all but one of the questionnaire items. When asked which cues they had used to solve the problem, the groups differed, χ^2 (6) = 21.42, $p < .005$. The questionnaire responses were similar to the results from test trials except for *S*s in the omit condition. The test-trial data from the omit condition showed only .404 *S*s who learned both cues, whereas in answering the questionnaire, .592 claimed to have used both cues. (This proportion was only slightly less than the 62% of *S*s in the consistent condition who reported they learned both cues.)

TABLE 1

ACQUISITION OF THE TWO-CUE PROBLEM AS A FUNCTION OF BLANK-TRIAL CONDITIONS

Statistic	Blank-trial condition			
	Consistent	Conflict	Omit	Single-cue
No. of *S*s	118	131	186	59
Proportion of nonsolvers	.076	.076	.108	.153
Mean total errors	3.483	3.229	4.925	4.610
Mean trial of last error	8.97	9.32	10.75	11.83

TABLE 2

TEST-PHASE DATA: DISTRIBUTION OF SOLUTION TYPES

Condition	No. of redundant relevant cues learned		
	0	1	2
Consistent	.161	.347	.492
Single-Cue	.119	.305	.576
Omit	.134	.516	.350
Conflict	.191	.573	.237

More detailed analysis showed how *S*s' understanding of the experimental situation affected responses. In the conflict condition, the two redundant cues were in opposition on blank trials. Of *S*s in that group who used both cues, .566 reported that they believed the blank trials were irrelevant and ignored them. Of one-cue solvers, only .287 responded the same way. Generally speaking, except for two-cue solvers in the conflict situation, a majority of all groups believed that their blank-trial responses were correct.

Approximately 30% of *S*s in each condition reported that they ignored the blank trials, while 55%–60% reported they believed that their blank-trial responses were correct. A large minority of *S*s in each condition (39%–47%) reported that they modified their H set in some way as a result of the stimuli they viewed on the blank trials. While those distributions of responses did not differ among the four conditions, they all contradict the blank-trial assumption that the H set is not modified.

Discussion

Both the test-phase data and the questionnaire data support the major conclusion that the nature of the stimulus viewed on the blank trials may indeed influence the learning process as measured in this experiment. Blank-trial probes apparently cannot be used to evaluate H tracking in the RRC problem. Contrary to the predictions of the blank-trial assumption, *S* may encode and process information (i.e., modify his H set) on all trials, not only on feedback trials. The processing can become apparent, however, only if *S* holds more than one H in his set and the blank-trial stimuli contain information not given in feedback-trial stimuli. In the present experiment the majority of *S*s in the consistent and single-cue conditions solved the problem by learning about both relevant cues. On the other hand, the omit and conflict conditions, where blank trials presented new and informative stimulus configurations, resulted in a restriction of the amount learned. This is the first demonstration in a CI problem that noncontingent presentations of blank trials may affect learning.

The reduction of two-cue solutions in the omit and conflict conditions is predicted from Restle (1962) and can be derived from the basic H-testing model of Trabasso and Bower (1968) with the added assumption that on blank trials *S* will eliminate any Hs from his set that are inconsistent with his response. That is, on blank trials *S* processes Hs in his H set exactly as he does on positive-feedback trials. This last assumption follows statements by Spence (1970), and to some extent Levine et al. (1964) and Richter and Levine (1965), that noncontingent-right and blank trials were equivalent, but departs from Levine (especially 1970) with respect to his blank-trial assumption. The blank-trial assumption states that although *S*s consider blank feedback as equivalent to being told their response was correct, the composition of the H set is not changed as a result of the blank feedback as it is on positive-feedback trials. Levine's and our assumptions make the same predictions in the following cases: when *S*'s H set contains only one H, when the blank-trial stimuli contain no new information that is not available on feedback trials, and when blank trials are presented contingent upon *S*'s response. In the first case, with only one H in the set, neither assumption would predict any effect due to blank trials. The two assumptions also make a common prediction (that blank trials will not affect learning) even though *S*s track several Hs both when the blank-trial stimuli are the same as the feedback-trial stimuli (as is true for Levine's, 1969, standard experimental setting and the present consistent condition) and when the blank-trial stimuli are a single cue (as Berger, 1970, and the present single-cue condition demonstrated in simple and RRC CI problems). The blank-trial assumption predicts no effect due to blank trials for both kinds of stimuli because *S* does not modify his H set on blank trials. The present assumption predicts no effects for slightly different reasons. When consistent stimuli are presented in a simple or RRC problem with relevant and irrelevant cues present and in conflict, an *S* tracking both relevant and irrelevant Hs would eliminate one or the other on a blank trial. If the relevant cue were dropped, *S*'s blank-trial responses would have indicated to *E* that *S* was tracking only the relevant cue in his sample, while in the other instance *S*'s responses would have indicated he was tracking only the irrelevant cue. The blank trials do not induce additional errors into the protocols in either case. Nor do the consistent blank trials in a RRC problem affect the distribution of solution types since an *S* tracking both RRCs would either maintain both or eliminate both on a given blank trial. The single-cue-condition blank trials have no effect because a single cue presented in isolation cannot be inconsistent with any Hs in *S*'s H set and so would never cause him to eliminate any cues from his sample.

In those H-testing problems where blank trials were presented contingent upon the correctness or incorrectness of *S*'s response, both assumptions yield the same predictions since blank-trial stimuli come from the same population as feedback-trial stimuli. Both correctly predict the equivalence of the right vs. wrong to the wrong vs. blank conditions as is found, but both also incorrectly predict that *S*s in the right vs. blank condition should not solve the problem. Other factors (e.g., contrast, information, etc.) have to be invoked to account for the learning that does take place.

The choice between the blank-trial assumption and the present assumption is moot since in the situations discussed thus far they both make the same predictions. The two assumptions can be empirically tested in the RRC conflict and omit conditions because they make differential predictions of the distribution of solution types that will be observed. Both conflict and omit blank trials should result in fewer *S*s learning both RRCs if the present assumption is correct, but should have no effect on two-cue solutions if the blank-trial assumption is correct. Any conflict-condition *S* who behaved according to the present assumption would have had to eliminate one of the relevant cues from his H set if both were present. Any *S*s who had sampled only one of the relevant cues would, of course, not notice the conflict and would proceed to solve the problem or eliminate the relevant cue and make a subsequent error. No *S* should learn both cues. In fact, this procedure reduced the proportion of two-cue solvers by about 30%. The problem, for the present assumption, becomes why 30% of the conflict-condition *S*s learned both cues. A partial explanation lies in the fact that a majority (18) of the conflict-condition two-cue solvers reported on the questionnaire that they ignored the blank trials; that is, at least some of the two-cue solvers believed that nonfeedback meant no information rather than a confirmation of their response, in support of the blank-trial assumption. It is unclear at present why the other 12 two-cue solvers learned both cues if, as they too reported, they did process the blank-trial stimuli. For the blank-trial assumption, the questions raised by these data are why 30% fewer *S*s solved the problem learning both cues and why all *S*s did not report they ignored the blank trials.

In the omit condition, too, there should be a reduction in the number of two-cue solvers if blank trials and positive feedback are processed the same, but not if blank trials are not processed. If *S* holds both relevant cues in his sample, he may discard the one that is present on the blank trial when it is in conflict with an irrelevant cue in his sample. The *S* would proceed to solve the problem using the remaining relevant cue in his sample. The reduction of two-cue solvers that should result would not be as marked as in the conflict condition because it would depend on *S*'s holding at least one irrelevant cue in his sample and discarding the relevant cue on the blank trial. If *S* were to eliminate the irrelevant cue before or on the blank trial, he would of course solve the problem knowing both relevant cues. In both this and the conflict condition, an *S* holding both RRCs in his H set would keep *both* if he did not eliminate Hs on blank trials.

There is an ambiguity in the omit-condition results because of the inconsistency regarding the number of two-cue solvers between the test-phase and questionnaire results, which may be due to *S*s' noticing the deleted cues but not learning the correct response assignments to them both (cf. Trabasso & Bower, 1968, pp. 85–86). These questionnaire responses are in accord with the blank-trial assumption.

Treating blanks as positive feedback would also result in a slightly increased error rate in the omit condition as was observed. The single-cue-condition mean errors should not have differed from the mean of the consistent condition, however, since relevant cues were not discarded. It is possible that deleting cues on the blank trials disoriented and confused *S*s (at least early in training) in both the omit and single-cue conditions. Such a disruption would not have affected the observed distributions of solution types, but could cause *S* to make an extra error. This notion also raises the possibility that if more detailed instructions or pretraining on the deleted blank-trial stimuli had been given in the present experiment, single-cue stimuli might have been found to be a valid tool for measuring H-testing behavior in RRC CI problems as Berger (1970) showed for a simple CI problem. Single-cue stimuli provide no information to the *S* who eliminates Hs on blank trials, but they can tell *E* which Hs *S* is tracking.

When new information is provided on the blank trials, either in the form of new stimulus configurations or in the contrast with feedback on other trials or in the information value of the feedback (Buchwald, 1962; Spence, 1970, 1972), blank trials have been found to affect learning. According to Levine's formulation of the blank-trial law for the simple CI situation, blank trials should not affect learning.

Two aspects of the present results that are not consistent with the present assumption that Hs are processed on blank trials exactly as on positive-feedback trials are the 29%–37% of *S*s in each condition who reported they ignored the blank trials and the conflict-condition *S*s who learned both RRCs although they reported they did not ignore the blank trials. This is in accord with Spence's (1970) finding that while the great majority of *S*s treat blank feedback as equivalent to positive feedback (or negative feedback in right vs. blank) a minority treat blank feedback as conveying no information. It is clear that either due to individual differences, training histories, etc., *S*s will be found who adopt different strategies in the problem. In this experiment there was minimal manipulation of *S*'s sampling strategies or the informational value of blank trials either by pretraining or instructions; yet almost all the data indicate the majority of *S*s both tracked several Hs and treated blank and positive feedback as equivalent in every respect.

As Levine (1970) stated with reference to the working H vs. majority rule alternatives, "the evidence suggests that subjects use either rule and it is the task for future research to determine when each occurs [p. 404]." The present experimental results can only be added as a caveat that the situation be set up so as to allow a behavioral differentiation of the alternative styles and strategies that *S*s may adopt.

REFERENCES

Berger, D. E. *Direct measures of information processing in concept identification.* (Doctoral dissertation, University of California, Los Angeles) Ann Arbor, Mich.: University Microfilms, 1970, No. 70-19828.

Bourne, L. E., Jr., Guy, D. E., & Wadsworth, N. Verbal-reinforcement combinations and the relative frequency of informative feedback in a card-sorting task. *Journal of Experimental Psychology*, 1967, **73**, 220–226.

Buchwald, A. M. Extinction after acquisition under different verbal reinforcement combinations. *Journal of Experimental Psychology*, 1959, **57**, 43–48.

Buchwald, A. M. Variations in the apparent effects of "right" and "wrong" on subsequent behavior. *Journal of Verbal Learning and Verbal Behavior*, 1962, **1**, 71–78.

Buss, A. H., & Buss, E. H. The effect of verbal reinforcement combinations on conceptual learning. *Journal of Experimental Psychology*, 1956, **52**, 283–287.

Izawa, C. Function of test trials in paired-associate learning. *Journal of Experimental Psychology*, 1967, **75**, 194–209.

Levine, M. Hypothesis behavior by humans during discrimination learning. *Journal of Experimental Psychology*, 1966, **71**, 331–338.

Levine, M. Neo-noncontinuity theory. In G. Bower & J. T. Spence (Eds.), *The psychology of learning and motivation.* Vol. 3. New York: Academic Press, 1969.

Levine, M. Human discrimination learning: The subset-sampling assumption. *Psychological Bulletin*, 1970, **74**, 397–404.

Levine, M., Leitenberg, H., & Richter, M. The blank trials law: The equivalence of positive reinforcement and nonreinforcement. *Psychological Review*, 1964, **71**, 94–103.

Moore, J. W., & Halpern, J. Three-stimulus two-choice auditory discrimination learning with blank trials. *Journal of Experimental Psychology*, 1967, **73**, 241–246.

Restle, F. The selection of strategies in cue learning. *Psychological Review*, 1962, **69**, 329–343.

Richter, M. L., & Levine, M. Probability learning and the blank trials law. *Psychonomic Science*, 1965, **2**, 379–380.

Rimm, D., Roesch, R., Perry, R., & Peebles, C. Effects of blank versus noninformative feedback and "right" and "wrong" on response repetition in paired-associate learning. *Journal of Experimental Psychology*, 1971, **88**, 26–30.

Slaymaker, F. L. Blank trial effects in concept identification. *Journal of Experimental Psychology*, 1972, **92**, 49–52.

Spence, J. T. Verbal reinforcement combinations and concept-identification learning: The role of nonreinforcement. *Journal of Experimental Psychology*, 1970, **85**, 321–329.

Spence, J. T. Effects of blank versus noninformative feedback and "right" and "wrong" on response repetition in paired-associate learning: A reanalysis and reinterpretation. *Journal of Experimental Psychology*, 1972, **94**, 146–148.

Trabasso, T., & Bower, G. H. *Attention in learning: Theory and research.* New York: Wiley, 1968.

(Received July 17, 1972)

Journal of Experimental Psychology
1973, Vol. 98, No. 2, 375–378

S–R COMPATIBILITY, RESPONSE DISCRIMINABILITY, AND RESPONSE CODES IN CHOICE REACTION TIME[1]

HARVEY G. SHULMAN[2] AND ALAN McCONKIE

Ohio State University

The effects of S–R compatibility and response discriminability on two-choice reaction times were studied by manipulating both S–R mappings and the ensemble of required responses. The effects of these two variables were equal in magnitude and almost perfectly additive. If S–R mappings are coded in terms of spatial transformation rules, as other evidence suggests, then the present results implied that interresponse relationships are represented by motor codes.

In a series of experiments beginning in 1953, Fitts demonstrated the importance of compatibility relationships for the rate of information transmission in speeded task performance (Fitts & Seeger, 1953). He found that the interaction of stimulus and response ensembles was at least as important as either of the two taken independently. He called this phenomenon S–R compatibility, and operationally defined as compatible any stimulus–response pairing which yields a high rate of information transmission when compared to other possible mappings of the same stimuli and responses.

Corollary to the notion of S–R compatibility are S–S and R–R compatibility which Fitts defined as follows:

> Stimulus–stimulus (S–S) compatibility effects arise whenever two or more different stimulus codes or sensory inputs are employed simultaneously. . . . response–response compatibility effects arise whenever two or more separate response processes are carried on concurrently . . . [Fitts & Briggs, 1959, p. 94].[3]

Research on divided and selective attention as well as Garner's (1970) work on integral vs. separable stimulus dimensions has yielded insights concerning the basis of S–S compatibility, which may be considered along such dimensions as discriminability and separableness. Response–response compatibility has been less well documented, however, and only recently has the effect of relationships between members of the same response ensemble come under investigation.

Fitts (Fitts & Briggs, 1959; Fitts & Seeger, 1953) began to expand on his operational definition of compatibility when he placed on it the theoretical interpretation of reflecting the complexity or number of recoding steps required in performance of a given task. The nature of these recoding operations was not specified by Fitts, but recent literature has begun to examine this issue. Experiments reported by Simon, Hinrichs, and Craft (1970) and Wallace (1971) were both concerned with the nature of the code which represents the mapping of stimuli into responses. Two possible forms of such a code are: (*a*) that it is spatial in the sense of coding the positional relationship between each element of the stimulus display and its assigned unit of the response device, or (*b*) that the code is motor in the sense of coding the relationship between each stimulus and its assigned response member (e.g., the right index finger). Both of the aforementioned studies found evidence favoring the spatial code, although Simon et al. also found results consistent with the hypothesis of an additional contribution of motor codes in their second experiment.

---

[1] This research was supported in part by Air Force Office of Scientific Research, Office of Aerospace Research, United States Air Force, Grant AFOSR-69-1669 to the Ohio State University Research Foundation.

[2] Requests for reprints should be sent to Harvey G. Shulman, Human Performance Center, 404B West 17th Avenue, Columbus, Ohio 43210.

[3] This report is the single readily available source of Fitts' discussion of the concepts of S–S and R–R compatibility. Copies may be obtained (at a cost of $3.00 each) by requesting Document AD714559 from the National Technical Information Service, Operations Division, Springfield, Virginia 22151. A related document by the same authors entitled "Skilled Performance, Part I" may also be obtained by requesting Document AD714558.

With regard to central codes, Posner and Keele (1972) have noted that the time to effect a correct response should be influenced by the tendency of a stimulus to activate competing motor programs, that is, programs which have a relatively large number of components in common with the program representing the correct response.

The present experiment was designed to test such a prediction and to further explore the nature of the codes involved in S–R and R–R relationships. Both S–R and R–R combinations were varied orthogonally in order to examine the possible interaction of their effects. Stimulus–response compatibility was varied by changing the spatial mapping rule in a two-choice RT task, which presumably affects the complexity of the spatial transformation rule. Response–response compatibility was varied by changing the composition of the particular response set used in the button-pressing task (Kornblum, 1965). In one discriminability condition the two index fingers were used, while in the other the index and middle fingers of the same hand were employed. This was assumed to change the number of control elements shared by the motor programs of the two responses, spatial mappings remaining constant.

METHOD

Subjects. Thirty-two right-handed males served as *S*s in response to an advertisement placed in the Ohio State University student newspaper. They were paid $1.25 per session. An additional four *S*s were dropped from the experiment after Session 1, either because their reaction times (RTs) were excessively slow or their error rates excessively high.

Apparatus. Stimuli were presented on a device consisting of a linear array of six Binaview displays. The display measured 27.94 cm. across, and it was placed 1.82 m. from *S*. Preprogrammed paper tape provided a control unit with information concerning character set and presentation intervals. The control unit also started an electronic clock simultaneously with each stimulus, and the depression of a response button closed a microswitch which stopped the clock. The response panel is described below.

Design. The experimental task was a two-choice RT task requiring button-press responses to visual stimuli. Two levels of S–R compatibility were factorially combined with two levels of response discriminability to define four treatment groups. Eight *S*s were assigned to each of these groups. In an initial experimental session, each *S* performed a sequence of 600 simple RT trials, including 150 catch trials on which no stimulus was presented. The stimulus on the remaining 450 trials was two zeros presented in the center of *S*'s display panel. The *S* responded to the onset of this stimulus by depressing a button on which he rested his right index finger. The mean RT for the final block of 100 simple RT trials was calculated for each *S*. Four groups of eight *S*s were stratified with respect to simple RT, and were then assigned to the four treatment combinations defined by S–R and R–R compatibility.

During a second experimental session lasting 45 min., each *S* responded to a random sequence of 600 two-choice RT trials in the experimental condition to which he was assigned. The two stimuli consisted of the letter X presented either at the extreme left or extreme right of the display panel, a distance measuring approximately 8° of visual angle. The two stimuli occurred with equal frequency. The response panel consisted of an array of six buttons arranged from left to right so that the first three fingers of each hand could comfortably be rested on them. The *S*s in one discriminability condition were instructed to use the buttons lying under their right and left index fingers, while in the other condition the right index and middle fingers were used. Simulus–response compatibility was varied by using either a direct or a crossed assignment of stimulus positions (left or right) to response members (leftmost or rightmost of whichever two fingers a given *S* was to use in responding).

Procedure. Each trial during the experimental session was initiated by the appearance of the word "ready" on the display for 2 sec., followed by the stimulus which remained visible until *S* depressed a response button. After *S*'s response, the ready signal for the next trial appeared and the cycle repeated. Following a run of 200 such trials, *S* was given a short break. Instructions emphasized the importance of responding quickly while maintaining an error rate of less than three or four errors in 50 trials.

RESULTS

Simple RT. The mean RT for each *S*'s final 100 simple RT trials was used as the criterion in assigning *S*s to treatment groups in the choice RT task. Mean simple RTs for the four choice RT groups are included in Table 1. Analysis of variance on simple RTs revealed no significant differences between groups. The standard errors of the group means ranged 37–55 msec. These RTs are somewhat long compared to most estimates of simple RT. Probably this is due to the relatively high proportion of catch trials included in the task and to the instructions given *S*s, which emphasized the avoidance of anticipation errors.

Choice RT. The sequence of 600 RT trials was broken down into 12 50-trial blocks and the mean RT for each *S* was calculated for each block. These means served as the basic data in subsequent analyses. Errors were not included in the calculation of mean RTs.

In each block of trials, separate means were calculated for responses made with the right index finger, and those made with either the left index or right middle fingers. In all four treatment groups the right index finger had been the designated response member on 300 of the 600 trials. Figure 1 shows mean RTs in each treatment group for right index responses only. The main effects of S–R compatibility and response discriminability were both significant, F $(1, 28) = 6.64$ and 6.53, respectively, $p < .025$. The interaction between these factors did not approach significance. Table 1 shows mean RTs over all trial blocks. The main effects of S–R compatibility and response discriminability on right index RTs are equal in magnitude (47 msec.) and almost perfectly additive.

The main effect of trial block on right index RTs was significant, $F\ (11, 308) = 4.57$, $p < .001$. This effect combined with S–R compatibility to produce the only significant interaction in the data, $F\ (11, 308) = 4.19$, $p < .001$. When the data from the final nine trial blocks were considered separately this interaction disappeared, $F\ (8, 224) = 1.25$. These results are due to the effects of practice which extended over the initial 150 trials only and which were greater in the conditions having a low degree of S–R compatibility.

TABLE 1

MEAN SIMPLE AND CHOICE REACTION TIMES (IN MSEC.) AND CHOICE ERROR RATES (IN PERCENTAGES)

Discriminability condition	Simple RT	Choice RT	Choice errors
Compatible			
Both index fingers	315	337	2.9
Index and middle fingers	321	379	2.6
Incompatible			
Both index fingers	313	379	5.5
Index and middle fingers	318	430	5.5

Note. Mean RTs are for right index finger responses only. Error rates are based on all responses.

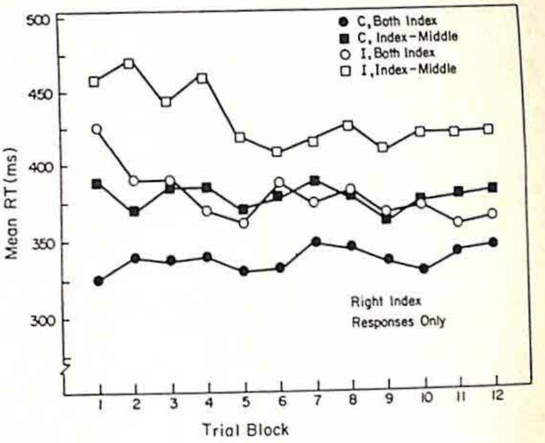


FIG. 1. Mean reaction time as a function of practice, plotted separately for each combination of S-R compatibility (C = compatible, I = incompatible) and response discriminability.

Mean RTs for the second finger (left index or right middle) used in each group were analyzed in the same way as data for the right index finger. These data show the same basic effects of S–R compatibility, response discriminability, and practice as the right index data. Thus, the main effects of S–R compatibility, $F\ (1, 28) = 11.60$, $p < .01$; response discriminability, $F\ (1, 28) = 4.38$, $p < .05$; and trial block, F $(11, 308) = 3.64$, $p < .01$, were significant. Both S–R compatibility and response discriminability interacted with trial block in the second finger RT data. There was a significant S–R Compatibility × Trial Block interaction, $F\ (11, 308) = 2.69$, $p < .01$, and a significant Response Discriminability × Trial Block interaction, $F\ (11, 308) = 1.92$, $p < .05$.

Error data. Table 1 includes error rates for each treatment group. These percentages represent the errors made with either response member in each group. An analysis of variance on these data revealed one significant effect, S–R compatibility, $F\ (1, 28) = 9.15$, $p < .01$. The effect of a decrease in S–R compatibility in this experiment, then, was to increase errors and decrease speed.

DISCUSSION

The most important results in this experiment were the demonstration of response discriminability effects, and the additivity of the effects of S–R compatibility and response discriminability. In order to interpret these facts, it is useful to adopt the logic of Sternberg's (1969) additive factors analysis. The power of this analysis is contained in the assumption that when two variables have additive effects they may be considered to influence separate and independent stages in the sequence of cognitive events which contribute to RT. Conversely, two variables which interact are considered to influence a single cognitive process.

In the present data the effects of S–R and R–R compatibility are additive and hence affect independent processes. Evidence presented by Simon et al. (1970) and Wallace (1971) indicates that the effects of S–R compatibility on RT are due to changes in the complexity of a spatial transformation rule used by *S* to relate stimulus location to the point in space at which a response must be made. It is suggested that the effects of R–R compatibility observed in the present data may be mediated by changes in the relationship between motor codes representing the movements required of *S* in making a response. The spatial code relating stimuli to response positions and the motor codes relating response movements to each other represent distinct constructs whose operations are independent; hence the additivity of the effects of S–R and R–R compatibility.

Smith (1968) proposed that four major processing stages are involved in choice RT performance. The first two were concerned with the processing and identification of the stimulus. Stage 3 involved the selection of a response, and Stage 4 the execution of the motor program representing that response. The effects of S–R compatibility would logically seem to be within Stage 3 where stimulus information is decoded into response information. The effects of response discriminability or R–R compatibility are more difficult to localize, either Stage 3 or Stage 4 being logical possibilities. The additivity of S–R and R–R compatibility effects is consistent with either possibility, since the two forms of compatibility might affect different stages in Smith's schema or component processes within a single stage.

REFERENCES

FITTS, P. M., & BRIGGS, G. E. Skilled performance: II. Task variables and performance. (Final Rep., Contract AF 41 657-70) Columbus: Ohio State University Research Foundation, March 1959.

FITTS, P. M., & SEEGER, C. M. S–R compatibility: Spatial characteristics of stimulus and response codes. *Journal of Experimental Psychology*, 1953, **46**, 199–210.

GARNER, W. R. The stimulus in information processing. *American Psychologist*, 1970, **25**, 350–358.

KORNBLUM, S. Response competition and/or inhibition in two-choice reaction time. *Psychonomic Science*, 1965, **2**, 55–56.

POSNER, M. I., & KEELE, S. W. Skill learning. *Handbook of research on teaching*. Washington, D.C.: American Educational Research Association, 1972.

SIMON, J. R., HINRICHS, J. V., & CRAFT, J. L. Auditory S–R compatibility: Reaction time as a function of ear–hand correspondence and ear–response location correspondence. *Journal of Experimental Psychology*, 1970, **86**, 97–102.

SMITH, E. E. Choice reaction time: An analysis of the major theoretical positions. *Psychological Bulletin*, 1968, **69**, 77–110.

STERNBERG, S. The discovery of processing stages: Extensions of Donder's method. *Acta Psychologica*, 1969, **30**, 276–315.

WALLACE, R. J. S–R compatibility and the idea of a response code. *Journal of Experimental Psychology*, 1971, **88**, 354–360.

(Received July 20, 1972)

Journal of Experimental Psychology
1973, Vol. 98, No. 2, 379–386

SCANNING AND DECISION PROCESSES IN RECOGNITION MEMORY[1]

MICHAEL C. CORBALLIS[2] AND AVRUM MILLER
McGill University, Montreal, Quebec, Canada

The *S*s watched series of 65 letters and indicated as quickly as possible whether each letter was or was not included in a memorized list which contained 3, 6, 9, 12, or 15 letters. Decision time was a linear function of the number of letters in the list, and the slope of the *no* function was somewhat steeper than that of the *yes* function. Variations in stimulus and response frequency affected decision time only slightly. These results support the notion of a high-speed scanning process in recognition memory, even when the memorized lists exceed the memory span. It is suggested that this process is not one of search, but serves simply to prime the individual elements of the memorized list.

Sternberg (1966, 1969) reported that the time a person takes to decide whether or not a given probe item is included in a memorized list is a linear function of the number of items in the list. He inferred from this that *S* scans the list serially to search for the critical item. He also found that the slope of the scanning function was the same for *yes* decisions as for *no* decisions, suggesting that *S* scans the list right through, regardless of where the critical item might be located, instead of terminating his scan when a match is found; that is, the scan is *exhaustive* rather than *self-terminating*. To explain why this might be so, Sternberg (1969) suggested that if *S* were to examine the outcome of each serial comparison while scanning, his scanning rate would be slowed down and the average decision time increased; it is therefore more efficient to complete the entire scanning process before examining outcomes.

One difficulty with this theory is that it predicts that the time taken to make *yes* decisions should be independent of the serial position of the critical item in the list. Sternberg (1969) stated that serial position effects were indeed negligible in his experiments, but others have reported faster *yes* decisions when the critical item is in terminal positions than when it is in the middle of the list (e.g., Burrows & Okada, 1971; Corballis, 1967; Corballis, Kirby, & Miller, 1972; Kennedy & Hamilton, 1969; Kirsner & Craik, 1971; Morin, DeRosa, & Stultz, 1967). To accommodate such effects, an alternative to the exhaustive-scan theory has been proposed. This states that *S* has direct access to some internal representation of the probe and judges whether or not the item is in the memorized list by examining the "strength" of this representation in memory, much as though the task were one of signal detection (e.g., Baddeley & Ecob, 1970; Corballis, 1967; Norman & Wickelgren, 1969; Okada, 1971). Corballis et al. (1972) have outlined how such a theory might accommodate most of the basic findings, including Sternberg's and including serial position effects.

A possible limitation of the direct-access theory, however, is that it does imply a monotonic increasing function, but there is no compelling reason to expect linearity. But this raises the question of whether the function really *is* linear. Most investigators, including Sternberg, have used lists of up to only about six or seven items, so it is conceivable that the function is actually some more complex one that is merely approximated by linearity over an initial range.

The present experiment is an attempt to put linearity of the decision-time function to more stringent test by examining the function for list lengths of up to 15 items. Previous evidence on extended list lengths has produced somewhat conflicting information. Wingfield and Branca (1970) reported a fairly good linear function for memory lists of up to 12 letters, although there was some indication that the functions for *yes* and *no* decisions were slightly different. However, Ross (1970) found that decision time was better approximated by a logarithmic function of list length than by a linear function, even after extended practice. He used a fixed-set procedure, in which each *S* always operated on the same four lists of one, two, four, and eight items. The lists, consisting of letters and digits, were constructed somewhat unusually in that the eight-item list always included the items on the four-item list, and so on down. This may explain why the function was logarithmic rather than linear.

---

[1] This research was supported by Grant 9425-10 from the Defence Research Board of Canada.

[2] Requests for reprints should be sent to Michael C. Corballis, Department of Psychology, McGill University, P.O. Box 6070, Montreal 101, Quebec, Canada.

Since the longer lists in the present experiment exceeded the normal memory span, we also used a fixed-set procedure. For each list length, *S* memorized a single list of letters and then received a number of trials for that list. Contrary to Ross (1970), Sternberg (1969) reported that the fixed-set procedure produced linear functions similar to those obtained under the varied-set procedure, in which *S* receives a new list to memorize before each probe. Our experiment was also designed so that the effects of stimulus and response probability could be assessed. In most experiments using Sternberg's paradigm, the effects of list length and stimulus probability have been confounded.

METHOD

Subjects. The *S*s were 20 university students, 5 male and 15 female, ranging in age 17–25 yr. They were paid for their services.

Apparatus and procedure. The basic procedure was to teach *S* a list of 3, 6, 9, 12, or 15 letters before presenting him with a series of 65 probe letters for that list. The lists were printed in large capital letters on cards and were learned to a criterion of two consectuive, correct reproductions. Recall was checked again after the probe trials and was perfect in all cases. Each *S* learned five lists, one of each length. The order in which the *S*s learned the lists was counterbalanced according to a Latin square, so that there were five equal groups of *S*s, each receiving a different order.

The first five trials of each series were treated as practice and were discarded from subsequent analysis. For the remaining 60 trials, the probe letters were selected in two different ways, and the *S*s were divided accordingly into two equal groups.

1. *Group R*: For this group the letters were chosen from the 26 letters of the alphabet in such a way that the number of positive probes always equaled the number of negative probes—30 in each case. That is, the a priori probabilities of *yes* and *no* responses were equal. However the probabilities that different letters would occur were unequal; for example, in the case of the three-letter list, each positive probe occurred 10 times while negative probes occurred an average of 1 7/23 times each.

2. *Group S*: Here, the frequencies of occurrence of the 26 letters were equated as nearly as possible; all letters occurred at least twice and eight of them occurred three times. However the a priori probabilities of *yes* and *no* responses were unequal; the number of positive probes ranged from 7 out of 60 for the 3-letter list to 35 out of 60 for the 15-letter list.

For both groups, positive probes were distributed as evenly as possible among the different serial positions, and in all other respects the sequences of probe letters were pseudorandomized to avoid obvious patterns or runs.

After *S* had learned a list, the 65 probe letters were projected on the screen of a digital display unit placed about 2 m. from where *S* was seated. The letters were in uppercase, 3.53 cm. high, and appeared white on the black screen. A warning tone sounded .5 sec. before the onset of each letter, which itself appeared for .5 sec. There was an interval of 3.5 sec. for *S* to make his response, followed by the warning tone for the next trial.

The *S* indicated his decision by pressing one of two buttons, placed 5 cm. apart on a response box. Between trials, *S* rested the index finger of his preferred hand on a rest position midway between the buttons. For half of the *S*s the *yes* button was to the side of the preferred hand and the *no* button to the other side; for the other half it was the other way about. A timer started synchronously with the onset of each letter and stopped either when *S* pressed a button or, if he failed to do so, when the warning tone for the next trial sounded. A printout device automatically recorded the response time and which of the two buttons *S* had pressed on each trial.

RESULTS AND DISCUSSION

Errors. Machine errors occurred on .85% of all trials. These trials were discarded from all further analysis.

The percentages of misses and false positives at each list length are shown in Table 1, averaged separately for Group R and

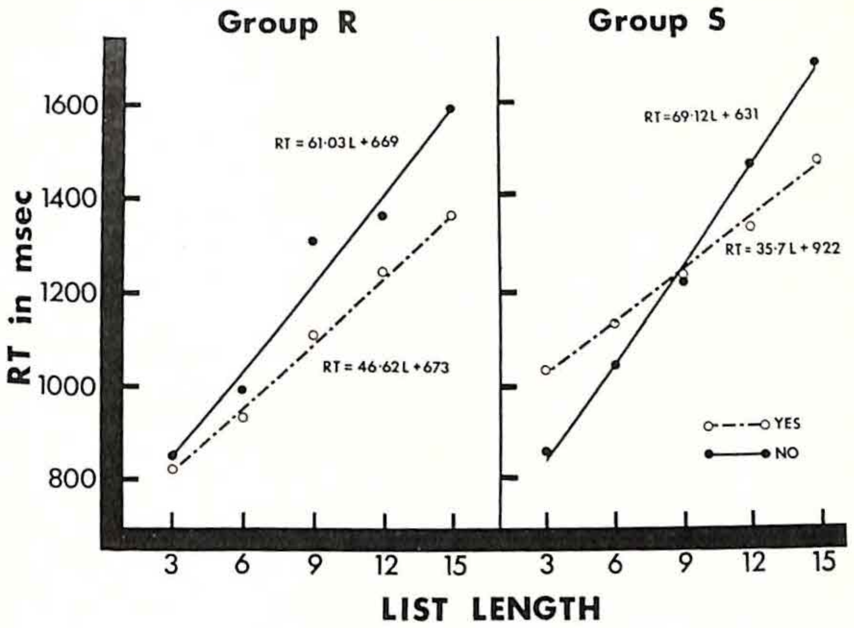


FIG. 1. Mean response times (RTs) for correct *yes* and *no* decisions at each list length, shown separately for Group R and Group S. (For group R, the a priori probabilities of *yes* and *no* responses were equal and the probabilities that differet letters would occur were unequal. For Groups S, the response probabilities were unequal and the stimulus probabilities approximately equal.)

Group S. The errors are too few (at least for the shorter lists) and individual differences too great for detailed inferences to be drawn, but it is at least clear that errors increased with list length and that misses outnumbered false positives. There is also some suggestion from a comparison between groups that the relative percentage of misses and false alarms varied inversely with the relative a priori probabilities of *yes* and *no* responses.

TABLE 1

MEAN PERCENTAGE OF ERRORS FOR EACH GROUP AT EACH LIST LENGTH

Groups[a]	Error type	List length				
		3	6	9	12	15
R	Misses	.67	1.67	7.67	10.00	13.00
	False positives	.67	.33	3.00	2.00	4.00
S	Misses	.71	5.00	2.14	7.83	4.57
	False positives	.19	.43	.38	1.17	4.00

[a] For Group R, the a priori probabilities of *yes* and *no* responses were equal and the probabilities that different letters would occur were unequal. For Group S, the response probabilities were unequal and the stimulus probabilities approximately equal.

The imbalance between misses and false alarms raises the question of bias in the analysis of response times, which was restricted to correct responses only (see following section). We shall argue from the results, however, that such biasing effects were probably small compared with the effects of other processes underlying response time.

RT and list length. Mean response times (RT) were computed for each correct decision (*yes* or *no*) at each list length for each *S*; *yes* RTs were averaged over all serial positions within a list. Means of these scores over *S*s are plotted separately for Group R and Group S in Figure 1.

Analysis of variance showed only that RT varied significantly with list length, F (4, 40) = 40.43, $p < .001$, and that there was a significant interaction between decisions (*yes* or *no*) and list length, F (4, 40) = 5.53, $p < .01$. Trend analysis of these effects revealed that a significant linear trend underlay the list-length function, F (1, 40) = 161.26, $p < .001$, and that this linear-trend component interacted significantly with decision, $F\ (1, 40) = 21.47$, $p < .001$; in both cases the deviation terms were insignificant ($F < 1$) and negligible. Thus, there is no basis for supposing the list-length functions to be other than linear, and the *no* function was reliably steeper than the *yes* function. Least-squares estimates of the slopes were 41.2 msec. per letter for the *yes* function and 65.1 msec. per letter for the *no* function.

Figure 1 indicates, however, that this difference in slope was somewhat greater in Group S than in Group R—least-squares estimates of the slopes are shown on the figure. The triple interaction between groups, decisions, and the linear-trend component of list length was significant only on a directional test, $F\ (1, 40) = 3.34$, $.05 < p < .10$, which is perhaps justified on the grounds that the direction of the interaction is predicted by the intergroup differences in stimulus and response frequencies.

First, let us consider the possible effects of *stimulus* frequency on the assumption that RT increases as stimulus frequency decreases (e.g., Bertelson & Tisseyre, 1966). For Group S, the average letter frequencies for positive and negative probes were essentially the same at each list length, so there are no grounds for anticipating a difference in slope between *yes* and *no* functions for this group. For Group R, however, the frequencies of positive probes decreased and those of negative probes increased, so that one might have expected the slope of the *yes* function to be greater than that of the *no* function. Figure 1 shows that exactly the reverse was the case, so that stimulus frequency effects cannot explain the slope differences. Nevertheless it is possible that stimulus frequency effects contributed to the fact that the difference in slope was less in Group R than in Group S.

Second, there is also evidence that RT increases as *response* frequency decreases (e.g., Dillon, 1966). In this case it is Group S, not Group R, which should be affected. One would predict for Group S that the slope of the *no* function should be steeper than that of the *yes* function, and this is in fact what happened. However response frequency cannot wholly explain the difference because it also occurred in Group R, although to a lesser extent. Nevertheless a response frequency effect could again have contributed to the difference between the two groups.

Thus stimulus frequency effects may have reduced the difference in slopes between *yes* and *no* functions in Group R, and response frequency effects may have increased it in Group S, in both cases contributing to the interaction between groups, decision, and list length. However, both classes of effects must have had only slight influence, at best, and we still need to explain why, overall, the *no* function was the steeper.

If the linear functions reflect a serial scanning process, then the fact that the *no* function was the steeper may mean that some *S*s were adopting a self-terminating scanning strategy—which would predict that the slope of the *yes* function would be about half that of the *no* function—while others were scanning exhaustively—which would predict equal slopes for both functions (Sternberg, 1969). To investigate this possibility, the slopes of the functions were computed, along with the squared correlations between list length and RT, for each *S*. Two *S*s were discarded from further consideration because the squared correlations were very low (.00 and .297, respectively), indicating that their RT–list-length functions were nonlinear. For the rest, the squared correlations ranged .689–.987, with a median of .880. For each of these 18 *S*s, the slope of the *yes* function was expressed as a ratio of the slope of the *no* function. The ratios, in increasing order, were .27, .32, .33, .50, .51, .52, .54, .59, .61, .65, .78, .81, .84, .88, .92, .99, and 1.24.

The distribution of ratios does appear to be bimodal, with one peak in the .51–.60 range and the other around .90. This gives some support to the notion that there were some *S*s who employed a self-terminating scanning strategy, at least for the most part, while others generally scanned exhaustively. To test this idea further, we divided the distribution into two: The 10 *S*s with ratios of .65 and below were pro-

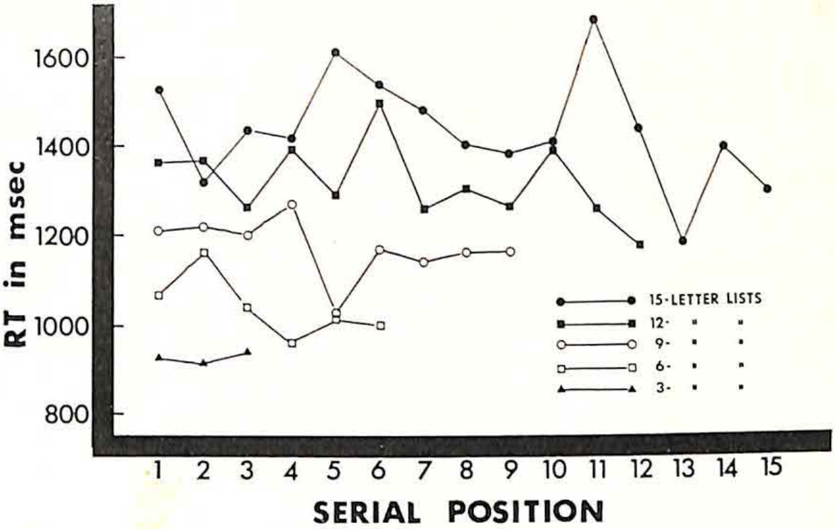


FIG. 2. Mean response times (RTs) for correct *yes* decisions for each serial position at each list length.

visionally labeled "self-terminating scanners," while the 8 *S*s with ratios of .78 and above were provisionally labeled "exhaustive scanners." The mean ratios were .48 and .92, respectively.

It is pertinent to compare the rates at which the two groups scanned. Sternberg (1969) argued that a self-terminating strategy would slow down the scanning rate, because *S* would have to evaluate each comparison during the scanning process. There is some earlier support for this; Clifton and Birenbaum (1970), using lists ranging from one to seven items, found that a few of their *S*s (3 out of 12) apparently used a self-terminating strategy and did indeed scan significantly more slowly than the others, who evidently scanned exhaustively. The mean slopes of the *no* functions were 46.7 and 31.0 msec. per letter, respectively. Our data show the same trend—the mean slopes were 78.0 msec. per letter for the self-terminating scanners and 61.1 for the exhaustive scanners—but the difference was not significant, t (16) = 1.41. However what is chiefly remarkable about these slopes is that *both* are considerably steeper than those reported by Clifton and Birenbaum. Even our so-called exhaustive scanners seem to have been scanning at only about half the normally reported rate.

An alternative explanation for these results is that *S*s may have sometimes scanned the lists more than once. Murdock (1971) has pointed out that this is more likely if the probe is negative than if it is positive, and it was certainly the case in our experiment that the slopes of the *yes* functions were closer than those of the *no* functions to the slopes typically reported in the literature—they were 37.4 and 55.3 msec. per letter, respectively, for the self-terminating and exhaustive scanners. Our labels may therefore be inappropriate; the groups may in fact be differentiated in terms of the relative numbers of scans following positive and negative probes, rather than in terms of the nature of the scanning process itself.

Serial position effects. The RTs for *yes* responses are plotted as a function of serial position in Figure 2, separately for each list length. Analyses of variance were also computed for each list length, with *S*s classified according to the order in which they received the different lists and according to whether they were in Group R or Group S. The effects of serial position were significant only in the case of the 6-letter lists, F (5, 50) = 4.31, $p < .01$, and of the 15-letter lists, F (14, 140) = 2.39, $p < .01$; however both become insignificant if the conservative degrees of freedom recommended by Greenhouse and Geisser (1959) are adopted. It is difficult to discern any systematic influences in these results; clearly, there are no straightforward primacy or recency effects.

In none of the above analyses did the classification variables interact significantly with serial position. They were therefore ignored in a subsequent set of analyses in which serial position effects were tested separately for the self-terminating and exhaustive scanners. There, only the six-letter lists revealed a significant serial position effect, and then only for the self-terminating scanners, $F\ (5,\ 45) = 4.83$, $p < .01$; again, the result is insignificant with reduced degrees of freedom.

These results raise further difficulties for the notion of self-terminating scanning. If the so-called self-terminating scanners had scanned from the beginning of the list, they should have shown clear primacy effects, but they did not. Clifton and Birenbaum (1970) faced a similar dilemma, since their three *S*s designated "self-terminating *S*s" either showed recency effects or no serial position effects at all. To maintain that scanning was self-terminating in these cases, one must suppose either that the *S*s reordered the lists before scanning or that scanning began at some random point in the list. Neither possibility carried much conviction. Again it is perhaps more plausible to suppose that the so-called self-terminating scanners were in fact scanning exhaustively, but were rescanning the lists more often following negative than following positive probes (cf. Murdock, 1971).

General Discussion

The most important finding was that the relation between response time and list length remained convincingly linear up to a list length of 15 items, more than double the maximum used by Sternberg (1966, 1969). Therefore, in spite of the arguments in favor of a direct-access strength theory (e.g., Baddeley & Ecob, 1970; Corballis et al., 1972), the evidence for a serial scanning process remains fairly compelling. And yet, on the other hand, we still face the dilemma raised in the introduction; many authors have reported strong serial position effects, which are clearly inconsistent with the exhaustive-scan theory.

There have been at least two previous attempts to develop a unified theory which can incorporate both serial position effects and the linear relation between decision time and list length. Following a suggestion by Atkinson, Holmgren, and Juola (1969), Murdock (1971) has proposed a model in which it is assumed that the list items are processed in parallel. The time to process a given item depends on its primary and recency in the list, as well as on the number of items still to be processed. By assuming a self-terminating process, and by making a number of other mathematical assumptions and approximations, Murdock shows that the function for negative probes is linear and that for positive probes nearly so, at least for list lengths up to five items. In addition, the predicted slopes of the positive and negative functions are about equal, and the *yes* decisions should show both primacy and recency effects. However the predictions break down for list lengths much longer than five items, unless it is assumed that scanning is exhaustive rather than self-terminating.

Burrows and Okada (1971) have proposed that an item in memory can be in a state of either high or low accessibility. If accessibility is high, the time taken to establish a match between the probe and the memorized item is shorter than the time taken to establish a mismatch, but if accessibility is low the match and mismatch times are equal. Assuming exhaustive scanning and a constant average probability, over different list lengths, that items will be in a state of high accessibility, the model then predicts equal linear functions for *yes* and *no* decisions, as well as serial position effects for *yes* decisions if items are in different states of accessibility. This theory is not without its difficulties, as its authors acknowledge. For example, the response times for *yes* decisions for a given list length could never be shorter than the *no* response times for shorter list lengths; however, this rule is contradicted both by the Burrows and Okada data and by those of Corballis et al. (972).

We propose another compromise theory here. We suggest first that the scanning process is not one of sequential *comparisons* at all, but is rather a sequential *priming* process which occurs prior to decision. This priming process would serve simply to activate the stored representations of the items—to give them "signal strength"—so they can then be distinguished along the decision axis from nonactivated ones. The decision would then be accomplished by direct access to the internal representation of the probe item, as proposed by such authors as Baddeley and Ecob (1970), Corballis et al. (1972), and Norman and Wickelgren (1969). The linear functions relating response times to list length can thus be attributed to the scanning process, while the serial position effects which often occur within a list can be ascribed to variations in the strength and criterion parameters which underlie the decision process. The effects of stimulus and response frequencies might also be attributed to the decision rather than to the scanning phase.

Priming may be necessary only when the items of a list have been serially organized in memory through rehearsal. A well-rehearsed sequence perhaps possesses an autonomy which prevents one from being directly aware of the elements of which it is composed, unless these elements are individually primed. This could be why the evidence for serial scanning is strongest in experiments where serial rehearsal is most likely to have occurred. One would also expect serial position effects to be minimal (though not necessarily absent) in such experiments, as indeed they are (Sternberg, 1969). On the other hand, priming may be unnecessary for items still in sensory storage, which itself may be regarded as a state of activation sufficient to mediate the decision process. For example, Okada (1971) reported strong recency effects and no evidence at all for serial scanning processes in a recognition-time experiment involving running memory, which would have precluded systematic rehearsal.

Besides reconciling otherwise conflicting evidence, the theory we have outlined overcomes certain logical difficulties in Sternberg's (1966, 1969) account. First, it is not very plausible to argue, as Sternberg did, that *S* makes a number of serial comparisons, each of which presumably has an outcome, while at the same time remaining oblivious to the outcomes. Thus *S* seems in one sense to be making comparisons, and in another sense not to be. Second, as Corballis et al. (1972) have pointed out, Sternberg's theory implies an infinite regress. The outcomes of the comparisons must themselves be stored as serial information for subsequent evaluation, which presumably requires another exhaustive scan of the outcomes. But then the outcomes of *that* scan must be stored serially and scanned—and so on.

However, there is at least one objection to our view; it is perhaps difficult to understand why *S* would not prime the memorized list *before* the probe appears. This would presumably lead to a more rapid decision, since it would remove the scanning time from the response time equation. One possible explanation, admittedly ad hoc, is that scanning is a rapid but fragile process whose priming effects would be easily disrupted by delay or by presentation of the probe. We have already suggested that *S*s may sometimes scan the list more than once, which implies that the priming may not always be effective. There may in fact be a trade-off between speed of scanning and strength of priming. The optimal strategy may therefore be to scan as rapidly as possible, but to delay scanning until after the probe has appeared so that the primed traces will still be sufficiently activated to mediate the decision.

REFERENCES

Atkinson, R. C., Holmgren, J. E., & Juola, J. R. Processing time as influenced by the number of elements in a visual display. *Perception & Psychophysics*, 1969, **6**, 321–326.

Baddeley, A. D., & Ecob, J. R. *Reaction time and short-term memory: A trace strength alternative to the high-speed exhaustive scanning hypothesis.* (Tech. Rep. No. 13) San Diego: University of California, Center for Human Information Processing, 1970.

Bertelson, P., & Tisseyre, F. Choice frequency of occurrence. *Nature*, 1966, **212**, 1069–1070.

Burrows, D., & Okada, R. Serial position effects in high-speed memory search. *Perception & Psychophysics*, 1971, **10**, 305–308.

Clifton, C., & Birenbaum, S. Effects of serial position and delay of probe in a memory scan task. *Journal of Experimental Psychology*, 1970, **86**, 69–76.

Corballis, M. C. Serial order in recognition and recall. *Journal of Experimental Psychology*, 1967, **74**, 99–105.

Corballis, M. C., Kirby, J., & Miller, A. Access to elements of a memorized list. *Journal of Experimental Psychology*, 1972, **94**, 185–190.

Dillon, P. J. Stimulus versus response decisions as determinants of the relative frequency effect in disjunctive reaction time performance. *Journal of Experimental Psychology*, 1966, **71**, 321–330.

Greenhouse, S. W., & Geisser, S. On methods in the analysis of profile data. *Psychometrika*, 1959, **24**, 94–112.

Kennedy, R. A., & Hamilton, D. Time to locate probe items in short lists of digits. *American Journal of Psychology*, 1969, **82**, 272–275.

Kirsner, K., & Craik, F. I. M. Naming and decision processes in short-term recognition memory. *Journal of Experimental Psychology*, 1971, **88**, 149–157.

MORIN, R. E., DEROSA, D. V., & STULTZ, V. Recognition memory and reaction time. *Acta Psychologica*, 1967, **27**, 298–305.

MURDOCK, B. B. A parallel-processing model for scanning. *Perception & Psychophysics*, 1971, **10**, 289–291.

NORMAN, D. A., & WICKELGREN, W. A. Strength theory of decision rules and latency in retrieval from short-term memory. *Journal of Mathematical Psychology*, 1969, **6**, 192–208.

OKADA, R. Decision latencies in short-term recognition memory. *Journal of Experimental Psychology*, 1971, **90**, 27–32.

ROSS, J. Extended practice within a single-character classification task. *Perception & Psychophysics*, 1970, **8**, 276–278.

STERNBERG, S. High-speed scanning in human memory. *Science*, 1966, **153**, 652–654.

STERNBERG, S. Memory-scanning: Mental processes revealed by reaction-time experiments. *American Scientist*, 1969, **57**, 421–457.

WINGFIELD, A., & BRANCA, A. A. Strategy in high-speed memory scanning. *Journal of Experimental Psychology*, 1970, **83**, 63–67.

(Received July 24, 1972)

Journal of Experimental Psychology
1973, Vol. 98, No. 2, 387–395

MEASUREMENT OF ORGANIZATIONAL PROCESSES WITHIN MEMORY STAGES[1]

DAVID W. MARTIN,[2] PAUL T. MARSTON,[3] AND RICHARD T. KELLY

New Mexico State University

In three experiments organization and memory load were investigated using a residual processing capacity methodology. Experiment I used compensatory tracking as a secondary task to show that lists which could be categorized required less processing in all stages of memory, encoding, retention, and retrieval, than lists which could not be easily categorized. This result was replicated using simple reaction time (RT) as a secondary task in Experiments II and III. In addition, processing requirements decreased during storage of lists which were ordered by categories at presentation, while no decrease was found for lists required organization. The RT was also found to be sensitive to list length.

A dual-task technique has recently been used to study information processing in memory (Johnston, Greenberg, Fisher, & Martin, 1970; Martin, 1970; Posner & Boies, 1971; Trumbo & Milone, 1971). While a number of names have been given to this technique, here it will be called residual processing capacity (RPC) methodology. The rationale behind RPC is that man has an upper limit to his capacity for processing information. For highly motivated *S*s, during short time intervals, this capacity can be considered fixed. When *S* is given a task which does not exceed his capacity, an amount of that capacity is left unused. The *S*'s performance on a secondary task should indicate how much RPC exists and conversely how much capacity is being expended on the primary task. The series of experiments reported here used the RPC methodology to investigate both memory load and organizational processes.

Several investigators have attempted to use RPC to localize the effects of organization to a particular stage of memory (Johnston, Wagstaff, & Griffith, 1972; Martin, 1970). The general procedure in these experiments is as follows. A verbal memory task is temporally divided into two stages: encoding, during presentation of a list of words, and retrieval, during recall of the words. These lists are presented to *S*s using one of three organizational schemes. For random lists each word in a list is a member of a different taxonomic category. In blocked and nonblocked lists several categories are represented by an equal number of words; in the former, the words from each category occur sequentially at presentation, and in the latter, the words are mixed at presentation.

This procedure permits a number of comparisons to be made. It has been known for some time that *S*s will organize lists and recall them in category clusters (Bousfield, 1953). Blocked lists do not require organization by *S* since they are presented in an organized manner. Nonblocked lists require organization but this is relatively easy since the words are members of only a few categories. If nonblocked lists are found to be organized when recalled by *S*s, and the RPC for these lists is less than for blocked lists during a given memory stage, then it can be inferred that the organization occurred during this stage. Random lists not only lack organization at presentation, but it is relatively difficult for *S*s to organize them since each word is a member of a different category. Hence, the RPC for these noncategorized lists can be compared to that for the categorized lists in order to determine in which stage the facilitatory effects of categorization occur.

---

[1] Experiment I was supported in part by the Air Force Office of Scientific Research, Office of Aerospace Research, U.S. Air Force under Grant AFOSR-69-1669 with the Ohio State University Research Foundation.

[2] Requests for reprints should be sent to David W. Martin, Department of Psychology, Box 3452, New Mexico State University, Las Cruces, New Mexico 88003.

[3] Now at University of Texas at Austin.

The results of two studies are of special importance. Martin (1970), in two experiments comparing blocked with nonblocked presentations using compensatory tracking as a secondary task, found no difference between conditions in either memory stage on the first presentation trial. However, over three trials using the same list, RPC was found to increase for blocked lists and stay about the same for nonblocked lists during the retrieval stage. This result indicated that blocked lists were easier to learn over trials due to ease of retrieval. Johnston, Wagstaff, and Griffith (1972) compared nonblocked and random lists using simple reaction time (RT) probes as a secondary task. Again there was no effect of organizational condition in either memory stage during the first trial. Over 12 trials, nonblocked lists allowed progressively more RPC than did random lists, particularly during retrieval.

The importance of determining the effects of organization on the first trial lies in the nonorganizational changes that can take place over trials. It is possible for *S* to encode all of the words on the first trial and recall them perfectly. If he does this he need not attend to the list during the second trial or during subsequent trials. Thus, the information processing requirements over trials can change regardless of organization. The clearest indication of organizational effects should occur during the first trial.

The series of experiments reported here attempted to more thoroughly examine first trial organizational effects. A possible reason for the previously reported lack of first trial organizational effects is *S*'s inability to efficiently use the organizational information available due to short trials during which little rehearsal is possible. Thus, in Experiment I a retention stage was provided in addition to encoding and retrieval stages. Also, three types of list organization were directly compared: blocked, nonblocked, and random.

EXPERIMENT I

Method

Subjects. Nine undergraduate and graduate student volunteers served as *S*s earning $1.25/hr plus incentive pay.

Tracking task. The secondary task consisted of visual compensatory tracking with a cyclical input and a positional control. A detailed description of this task can be found elsewhere (Martin, 1970). The dependent measure was integrated absolute tracking error in arbitrary voltage units. The *S*s practiced the tracking task alone, the verbal task alone, and then the tasks concurrently for five 40-min. training sessions prior to the test trials. Each *S* tracked continuously through each test trial. However, separate tracking scores were recorded during the 16 sec. in which eight words were being aurally presented (encoding), the 16 sec. when words were being retained (retention), and the 16 sec. of spoken free recall (retrieval).

Verbal task. A word pool was formed using the four most frequent associates to each of 36 category classifications (Cohen, Bousfield, & Whitmarsh, 1957). From this pool 18 eight-word blocked, nonblocked, and random lists were constructed, each word appearing under each condition. Three sets containing six lists from each condition were formed. Each set comprised the verbal material for three *S*s.

The word lists and all instructions to *S* were prerecorded on magnetic tape. The two category names were spoken prior to list presentation for blocked and nonblocked lists; "eight category" preceded random lists. This category instruction provided an organizational scheme to *S* so that the difference between conditions was one of organizational processing rather than the searching for a means of organization. Following this instruction the eight words were aurally presented at a 2-sec. rate. In blocked lists the four members of one category preceded those from the second. In nonblocked lists four members from each of two categories were presented, but no more than two words from the same category occurred in sequence. Random lists were composed of one word from each of eight categories. Following list presentation *S* repeated a three-digit number which was read to him. Pilot work had indicated that *S*s would often vocally rehearse the final three or four words in a list in the order that they had been presented. The three-digit number prevented this overt rehearsal of the final items for a second or two and apparently eliminated it, thereby enhancing the tendency to organize the material. After the number was repeated, a 16-sec. retention period occurred. It was terminated by a "recall" instruction. Sixteen seconds were allowed for spoken recall. Trials were separated by 15-sec. rest periods.

In order to encourage *S*s to devote their attention primarily to the verbal task, they received incentive pay in addition to $1.25/hr base pay. Incentive pay could be as much as $2/session, 80% determined by verbal performance and 20% by tracking.

Design. A within-*S* $3 \times 3 \times 6$ factorial design was used representing three levels of organization: blocked, nonblocked, random; three memory stages: encoding, retention, retrieval; and six different lists. In addition two control trials were presented during the test session in which *S* simply tracked with no words presented. These scores provided base-line tracking data.

Results and Discussion

The mean number of words correctly recalled per list were 7.9, 7.9, and 6.1 for blocked, nonblocked, and random conditions, respectively. A one-way analysis of variance showed this effect to be significant, $F\,(2, 16) = 77.2$, $p < .01$.

To determine whether *S*s did organize lists at recall, a clustering ratio (CR) was computed. This measure is simply the number of obtained category repetitions over the number expected by chance (see E [SCR] in Bousfield & Bousfield, 1966). A CR of 1.0 is chance performance. A $\mathrm{CR} > 1.0$ indicates that organization is taking place. In this case 2.0 is maximum. Apparently a very high degree of organization took place, since the CRs for both blocked and nonblocked lists were extremely high, 1.98 and 1.99, respectively.

Tracking scores are shown in Table 1. An analysis of variance on the means for each *S* showed list organization to be the only significant effect, $F\,(3, 24) = 15.7$, $p < .01$. A Newman-Keuls analysis ($p < .05$) showed the control (C) condition to be different from each experimental condition and the random (R) condition to be different from blocked (B) and nonblocked (NB) conditions ($\mathrm{C} < \mathrm{B} = \mathrm{NB} < \mathrm{R}$).

Thus, an effect of list organization was found in Experiment I. Even though fewer words were recalled per list, it was apparently more difficult to encode, retain, and retrieve random lists than either blocked or nonblocked lists. The major difference between this research and earlier studies which reported no organizational effect was the addition of a retention stage in the memory task. Encoding and retrieval stages may not allow the effects of list organization to become apparent in only one trial.

It seems somewhat surprising that no differences were found in secondary task performance for blocked and nonblocked conditions even though organization was required, and apparently occurred, in the latter. It is, of course, possible that words from lists are encoded irrespective of list organization at presentation, in which case blocked lists would also have to be reorganized. However, the superiority of recall performance for blocked lists which is usually found would indicate that nonblocked lists do require more processing (Calfee & Peterson, 1968). This processing should be apparent using an RPC methodology.

TABLE 1

TRACKING ERROR IN (VOLTS) AS A FUNCTION OF ORGANIZATION AND STAGE

Organization	Memory stage		
	Encoding	Retention	Retrieval
Blocked	222	223	232
Nonblocked	223	214	234
Random	240	243	260
Tracking control	178	182	183

Research to this point has been concerned with identifying a memory stage where organization occurs. Rather than having effects between stages, however, it is possible that organization could have within-stage effects. For example, organization might first have detrimental and then facilitatory effects during a particular stage. A comparison of stage means does not permit investigation of such interactions. Experiment II was designed to provide such information.

EXPERIMENT II

In order to examine within-stage effects, a simple RT secondary task was used in Experiment II. Simple RT has been shown to be a good index of RPC (Johnston, Griffith, & Wagstaff, 1972; Johnston, Wagstaff, & Griffith, 1972; Stanners, Meunier, & Headley, 1969). When used as a secondary task, RT offers the advantage of providing a fine-grained analysis of RPC. Posner and Boies (1971), for ex-

TABLE 2

MEAN WORDS RECALLED AND CLUSTERING RATIOS AS A FUNCTION OF ORGANIZATION

Organization	Words recalled			Clustering ratio		
	Four	Six	Eight	Four	Six	Eight
Blocked	3.8	5.2	6.9	1.84	1.72	1.84
Nonblocked	3.8	5.1	6.4	.52	1.32	1.38
Random	3.9	4.7	5.3	—	—	—

Note. The words four, six, and eight refer to four-word, six-word, and eight-word list conditions, respectively.

ample, used it as an indication of RPC during the 1.0 sec. between letter presentations in a letter comparison task. Tracking error, on the other hand, must be integrated over a finite period of time. Thus, measuring within-stage effects using a tracking secondary task is quite difficult. The first goal of Experiment II then, was the determination of effects within memory stages.

A second goal of Experiment II was to replicate the overall list organization results found in Experiment I using RT. A third purpose of the experiment was to more thoroughly validate the use of RT as an RPC measure by showing it is sensitive to general memory loading effects. This was accomplished by varying list length.

Method

Subjects. Thirty students served as *S*s in partial fulfillment of the requirements for an introductory psychology course. No incentive pay was given, although *S*s were told that the verbal task was more important, to be done at the expense of the RT task.

Verbal task. The verbal memory task was quite similar to Experiment I with the following exceptions: Word lists were presented visually with the category designations, and all words on a list were printed on a single slide. The slide was projected on a 60 × 60 cm. screen approximately 1.5 m. in front of *S*. This slide remained on during the encoding stage. A blank slide appeared during retention. The retrieval stage, during which the words were orally recalled, was initiated with the appearance of a slide with the word "recall" on it. Three list organizations were again chosen: blocked, nonblocked, and random. In this case category members were from frequently occurring items in the Battig and Montague (1969) norms. Four within-*S* conditions were used, the three list organizations plus a control condition in which the slides simply said "reaction time," and *S* could devote his entire capacity to the RT task.

In addition to the within-*S* conditions, the number of words comprising a list served as a between-*S*s variable. The 10 *S*s in each condition received either four, six, or eight words per list. The length of the memory stages was dependent on list length. At a 2-sec. rate the length of each stage was 8, 12, or 16 sec. The memory stages and rest periods were of equal length.

RT task. At some point within each stage of the memory task ($\frac{1}{8}$, $\frac{3}{8}$, $\frac{5}{8}$, or $\frac{7}{8}$ of the stage) a highly discriminable red jewel lamp was lighted at the bottom of the screen. The *S* was instructed to push a button which he held under his preferred thumb whenever this occurred. Latencies were automatically recorded to the nearest millisecond using a digital timer and printer. Within a particular stage, one of the intervals was randomly chosen for signal presentation with the constraint that each of the four possible intervals was presented five times to each *S* for each condition. Thus, each *S* in the experiment provided 240 RTs, representing three list organizations plus a control condition, three memory stages, four intervals within each stage, and five different trials. The medians of the five RTs were used as basic data.

Results and Discussion

The mean number of words correctly recalled and the CRs are shown in Table 2. It may be seen that as list length increased, the effects of list organization became more pronounced in the recall data. The CR scores indicate that *S*s recalled the blocked lists in a highly organized way. Some organization occurred in the six- and eight-word nonblocked lists. However, the four-word nonblocked lists actually showed organization below chance levels. Inspection of the words recalled indicates that *S*s often recalled the four-word lists in serial order. Since organization was deliberately minimized in the presentation of nonblocked lists, this result is not too surprising.

To eliminate inter-*S* variability due to base-level performance on the simple RT task, a single mean was determined for each *S* from his control condition. This number was subtracted from each of his experimental condition scores giving a deviation score. This procedure eliminates base-level differences between groups, thereby allowing assessment of uncontaminated effects due to number of words.

Figure 1 shows mean deviation RTs averaged across list lengths. An analysis of variance showed a highly significant effect

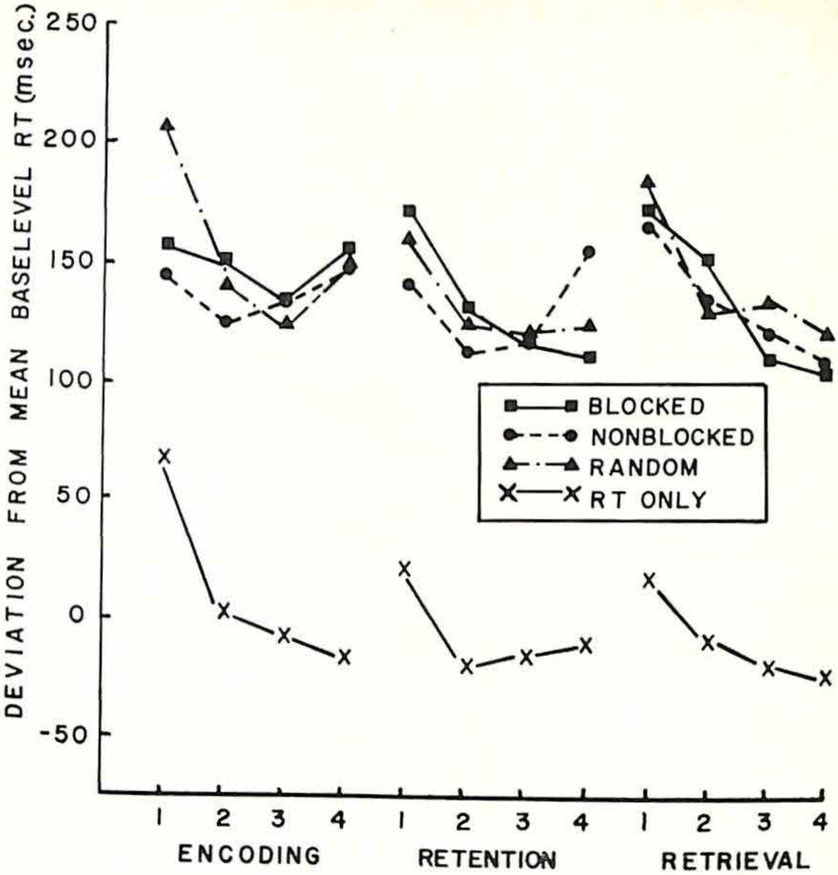


FIG. 1. Mean deviation reaction times (RTs) during Experiment II at each probe point for blocked lists, nonblocked lists, random lists, and RT only.

due to the four organizational conditions including the control condition, F (3, 81) = 241.65, $p < .01$. Since it is obvious that this effect was due largely to the difference between the control condition and all experimental conditions, a second analysis was conducted which excluded the control condition. While it is not apparent from the figure, this analysis showed that a small but significant source of variation was accounted for by list organization, F (2, 54) = 3.70, $p < .05$. A Newman-Keuls analysis ($p < .05$) showed the random list mean to be larger than the nonblocked mean. The difference, however, was small, 9.4 msec. Of more interest are the following interactions: Organization × Stages × Intervals, F (12, 324) = 2.62, $p < .01$; Organization × Intervals, F (6, 162) = 5.52, $p < .01$; and Organization × Stages × List Length, F (8, 108) = 2.37, $p < .05$. Inspection of Figure 1 indicates that these effects are due in part to the large RT for the random condition at the first probe point, particularly in the four-word condition. It is possible that *S* was attempting to find a basis for organizing the list here. He was given the two category names as a basis for organization in the other conditions. A second interesting contributor to the first of these interactions is the general downward trend of RT for blocked lists during retention and the upward trend for nonblocked lists. Perhaps blocked lists become easier to rehearse as retention continues, while nonblocked lists become more difficult due to organization.

The small size of the organizational effects could be due to a general lack of sensitivity using RT to assess RPC. In order to determine sensitivity one can examine the effect of a general loading variable such as

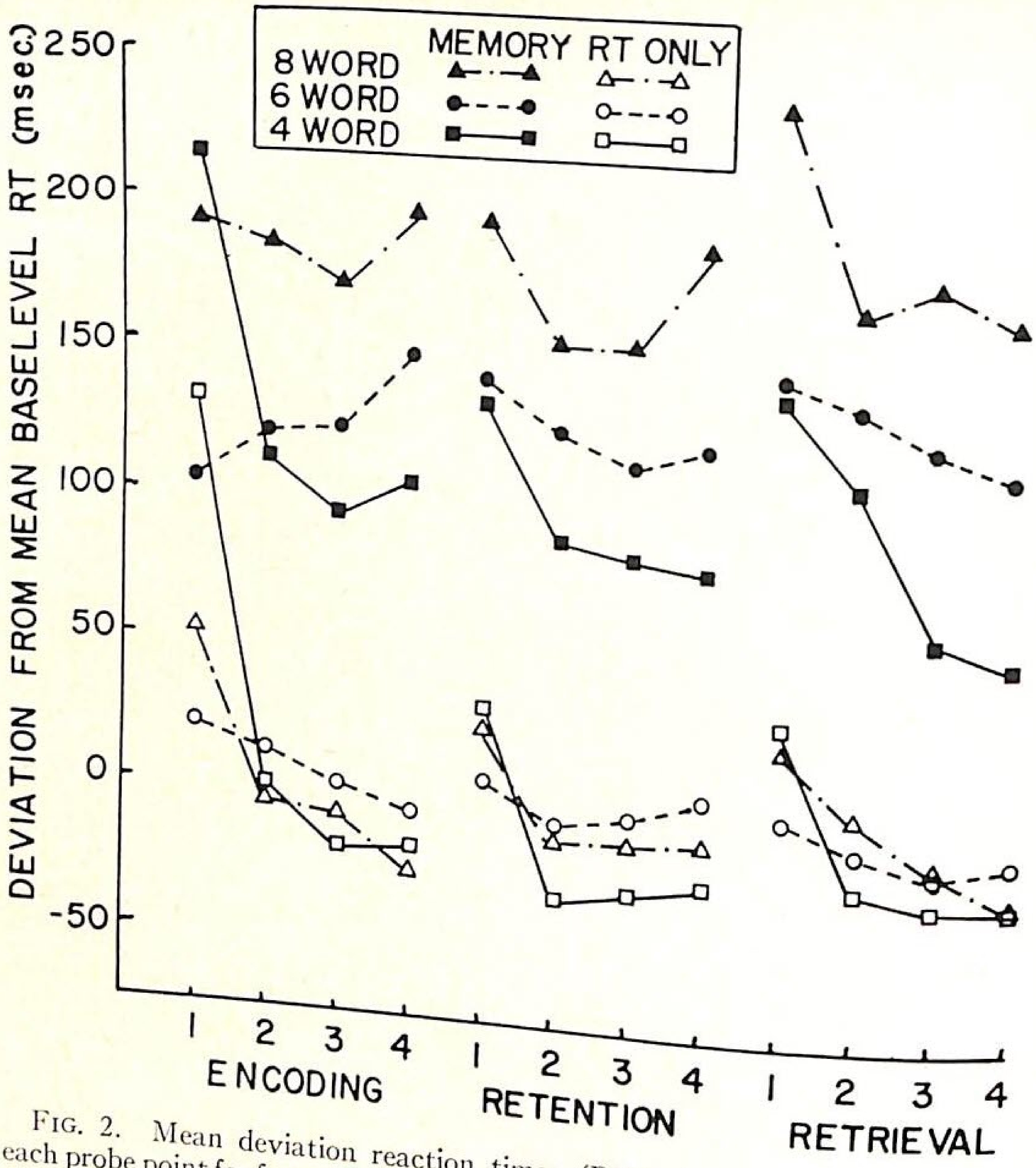


FIG. 2. Mean deviation reaction times (RTs) during Experiment II at each probe point for four-, six-, and eight-word lists, and corresponding RT-only conditions.

list length. The main effect of list length reached a high level of significance, F (2, 27) = 7.88, $p < .01$. The size and consistency of this effect can be seen in Figure 2. The single point which is out of order is the source of several significant tests: stages, F (2, 54) = 3.58, $p < .05$; List Length × Stages, F (4, 54) = 3.24, $p < .05$; and List Length × Stages × Intervals, F (12, 162) = 3.72, $p < .01$. Since this point is also displaced in the control condition, it is apparent that the effect is not caused by the memory task. In the four-word condition less than 1 sec. passed between the slide change and the first RT signal. It is possible that *S* was not expecting the signal so soon after the start of a trial. This effect is apparent only in the four-word condition.

Inspection of Figures 1 and 2 indicates that the significant intervals effect, F (3, 81) = 37.90, $p < .01$, is due to the general decrease in RT during each stage. Since this decrease was also apparent in the control condition, it may be due to an expectancy effect. Even though *S* was not informed that a signal would appear in each stage, this became apparent fairly early in a session. Once he perceived this, the probability of a signal occurring at each interval given that it had not occurred yet increased throughout a stage. Hence the common finding that the more likely an event, the more quickly the response to that event.

The Stages × Intervals interaction, F (6, 162) = 3.67, $p < .01$, indicates that the drop in RT from Interval 1 to Interval 4 was larger for retrieval (63 msec.) than for encoding (20 msec.) or retention (27 msec.). In fact, the decrease for encoding was considerably less than the control condition, while the decrease for retrieval was considerably greater than the control condition. These data may be interpreted as a general increase, compared to the control condition, in RT during encoding and a general decrease during retrieval. This type of effect would be expected as *S* "loads" and "unloads" memory. However, caution must be exercised in this interpretation since the encoding result could be interpreted as a decrease in the expectancy effect with increased memory load.

Several important points are suggested by Experiment II. First the list length data indicate that RT is sensitive to memory loading. In addition, the data suggest a possible loading–unloading effect. Second, while the size of the effect was small, nevertheless, the general superiority of blocked and nonblocked lists to random lists found in Experiment I was replicated. Finally, a significant trend of increased RPC during retention of blocked lists and decreased RPC during storage of nonblocked lists was found.

EXPERIMENT III

The final experiment was conducted for several reasons. Some of the results reported in Experiment II were based on small differences. Replicating these results is important. In addition, several of the assumptions of the RPC model were violated in Experiment II. Perfect recall of words was not achieved, especially with eight-word lists. Thus, in the eight-word lists the primary task exceeded processing capacity by itself. An RPC measure may be meaningless when this happens, since presumably no RPC exists to perform the secondary task. The secondary task was, of course, performed in this experiment. When the primary task exceeds processing capacity in this way, the basic model is inappropriate. A possible alternative model when this assumption is violated is one in which performance on the secondary task is degraded by the primary task to a certain low level. When this level is reached *S* may refuse to further degrade it and instead will sacrifice performance of the primary task. The trade-off suggested by this model is similar to that found in speed vs. accuracy experiments in which *S* will emphasize speed or accuracy up to a point, but beyond this point it is very difficult to force him to increase errors to gain speed or to lose speed by decreasing errors (Fitts, 1966). Hence, one reason for selecting the six-word condition for replication in Experiment III rather than the eight-word condition was to avoid violating the assumption of perfect primary task performance and thus stay within the more basic model of attention.

TABLE 3

MEAN WORDS RECALLED AND CLUSTERING RATIOS AS A FUNCTION OF ORGANIZATION

Organization	Words recalled	Clustering ratio
Blocked	5.5	1.85
Nonblocked	5.4	1.10
Random	4.9	—

The six-word condition was chosen rather than the four-word condition since clustering scores indicated that very little organization was occurring in four-word lists. It would, of course, be impossible to find the locus of organization in this case. For these reasons the six-word condition in which the fewest assumptions were violated was replicated in Experiment III. As an added attempt to achieve near perfect performance on the memory task, a $20 prize was offered for the best performance. Performance scores were weighted 80% on words recalled, 20% on RT.

Method

Subjects. Ten students similar to those in Experiment II served as *S*s. A $20 prize was offered for the best performance, 80% based on word recall and 20% on RT.

Procedure. The procedure was identical to the six-word condition of Experiment II. The only exception to this was the addition of an incentive prize.

Results and Discussion

In order to allow a number of useful comparisons, the data for Experiment III were combined with the six-word condi-

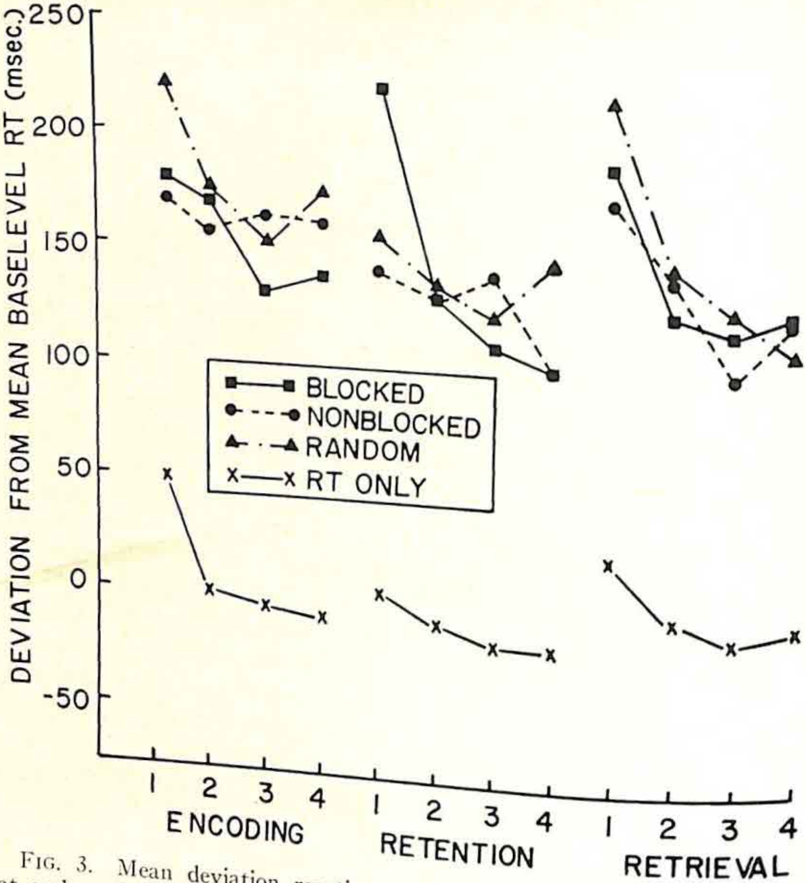


FIG. 3. Mean deviation reaction times (RTs) during Experiment III at each probe point for blocked lists, nonblocked lists, random lists, and RT only.

tions of Experiment II forming a between-*Ss* condition of incentive. Table 3 lists the mean number of words correctly recalled and the CRs. As can be seen in the table, the addition of an incentive weighted mainly on verbal performance caused an increase in average words recalled from 5.05 to 5.27, a significant increase, F (1, 398) = 9.96, $p < .01$. The RT also increased by an average of 20 msec. mainly during encoding and to some extent during retention. This main effect failed to reach significance, but since it is a between-*Ss* effect for only 10 *Ss* it may be an important trend.

The combined six-word results also completely replicate the significant effects of Experiment II except the stages effect (see Figure 3). The mean RT for random lists was significantly larger than blocked and nonblocked which were equivalent ($p < .01$). The Organization × Stages × Intervals interaction also reached significance, F (12, 216) = 2.63, $p < .01$, due largely to the significant ($p < .01$) drop of 86 msec. for blocked lists during storage.

GENERAL DISCUSSION

The three experiments taken together indicate that, unlike previously reported results which provided *Ss* little opportunity for rehearsal on the first trial, lists which can be organized by *S* into categories require less information processing during the entire memory trial than lists which cannot be easily organized. When comparing lists which can be easily organized but are not presented in an organized manner with those that are, an interaction is found within the retention stage of memory. Apparently, the latter lists require progressively less processing during rehearsal than the former. Thus, when a retention stage is provided in a memory task, lists which can be organized are apparently organized during retention, but when lists cannot be organized, the detrimental effects occur in all stages of a memory trial.

The experiments reported here also add further support for the use of an RPC methodology in assessing information processing requirements of a memory task when the assumptions of this model are met. In these experiments not only were overall differences found between loading conditions due both to list composition and list length, but within-memory stage effects were also found. It appears that an RPC methodology can be used to provide a profile of processing requirements during a memory trial.

REFERENCES

Battig, W. F., & Montague, W. E. Category norms for verbal items in 56 categories: A replication and extension of the Connecticut category norms. *Journal of Experimental Psychology Monograph*, 1969, **80** (3, Pt. 2).

Bousfield, A. K., & Bousfield, W. A. Measurement of clustering and of sequential constancies in repeated free recall. *Psychological Reports*, 1966, **19**, 935–942.

Bousfield, W. A. The occurrence of clustering in the recall of randomly arranged associates. *Journal of General Psychology*, 1953, **49**, 229–240.

Calfee, R. C., & Peterson, R. E. Effect of list organization on short-term probe recall. *Journal of Experimental Psychology*, 1968, **78**, 468–474.

Cohen, B. H., Bousfield, W. A., & Whitmarsh, G. A. *Cultural norms for verbal items in 43 categories.* (Tech. Rep. No. 22) Storrs: University of Connecticut, 1957.

Fitts, P. M. Cognitive aspects of information processing: III. Set for speed versus accuracy. *Journal of Experimental Psychology*, 1966, **71**, 849–857.

Johnston, W. A., Greenberg, S. N., Fisher, R. P., Martin, D. W. Divided attention: A vehicle for monitoring memory processes. *Journal of Experimental Psychology*, 1970, **83**, 164–171.

Johnston, W. A., Griffith, D., & Wagstaff, R. R. Speed, accuracy, and ease of recall. *Journal of Verbal Learning and Verbal Behavior*, 1972, **11**, 512–520.

Johnston, W. A., Wagstaff, R. R., & Griffith, D. Information processing analysis of verbal learning. *Journal of Experimental Psychology*, 1972, **96**, 307–314.

Martin, D. W. Residual processing capacity during verbal organization in memory. *Journal of Verbal Learning and Verbal Behavior*, 1970, **9**, 391–397.

Posner, M. I., & Boies, S. J. Components of attention. *Psychological Review*, 1971, **78**, 391–408.

Stanners, R. F., Meunier, G. F., & Headley, D. B. Reaction time as an index of rehearsal in short-term memory. *Journal of Experimental Psychology*, 1969, **82**, 566–570.

Trumbo, D., & Milone, F. Primary task performance as a function of encoding, retention, and recall in a secondary task. *Journal of Experimental Psychology*, 1971, **91**, 273–279.

(Received July 24, 1972)

Journal of Experimental Psychology
1973, Vol. 98, No. 2, 396–403

EFFECT OF SPEED OF LEARNING AND DEGREE OF LEARNING ON CUE SELECTION[1]

JACK RICHARDSON[2]

State University of New York at Binghamton

Consonant trigrams (CCCs) were used as stimuli in a paired-associate list and recall to the individual letters from the CCCs was tested following learning of the list. The efficiency of cue selection (the percentage of the recall to the individual letters which would be necessary for recall to the trigrams) was greater for *S*s who learned slower, for *S*s who had learned a list prior to the one tested for cue selection, and for difficult items. Efficiency decreased as degree of learning increased. However, the small decrease in efficiency with overlearning was *not* interpreted as due to relaxation of the selector mechanism.

Efficiency of cue selection has been specified as the degree to which an *S* learns the response to only one component of a compound stimulus (Richardson, 1971). It has been demonstrated that instructions and emphasis can increase the efficiency of cue selection and decrease the difficulty of learning a compound-stimulus list to the level of that obtained with single components as stimuli (Richardson, in press). The difference in difficulty of learning paired-associate lists with compounds and single components as stimuli was attributed to a lack of efficiency of cue selection with the compound stimuli.

One purpose of the present study was to determine the extent to which efficiency of cue selection contributes to the increased speed of learning produced by other variables. The three variables investigated were *S*'s speed of learning, item difficulty, and practice effects. It seemed that fast *S*s, easy items, and prior learning might all be related to increased efficiency.

The second purpose of the present study was to further investigate the effect of degree of learning on the efficiency of cue selection. James and Greeno (1967) found increased learning to the nonselected components only after the list was mastered, and this increased cue effectiveness with overlearning has been confirmed (Martin, 1971; Wichawut & Martin, 1970). However, Houston (1967) and Lovelace and Blass (1968) failed to find that overlearning produced more increase in the cue effectiveness of the nonselected components than of the selected component. Presentation of a response paired with a compound stimulus after learning the response to one component of the compound did not produce learning to the additional components in a paired-associate task (Richardson & Stanton, in press), but learning to additional components did occur in a discrimination task (Siegel & Stevenson, 1966). Berry, Joubert, and Baumeister (1971) found an increase in the number of two-letter solutions starting relatively early in learning with retardates as *S*s. Almost every possible empirical outcome has been obtained in studies varying degree of learning, and there is little to suggest what produces the different results. The present article varies degree of learning over a wide range and presents a detailed analysis of the variation in cue selection.

In Experiment I of the present article all *S*s learned a vowel–word list before learning a second list with compounds as stimuli and recalling to the components of the compounds. It seemed possible that learning a list with single components as stimuli might encourage selection from a compound stimulus list so, in Experiment II, learning of the vowel–word list occurred subsequent to learning and recall of the compound list. In Experiment III the types of stimuli in a list learned prior to compound-list learning were varied. Degree of learning of the compound list was varied in all three experiments.

---

[1] This research was supported by National Science Foundation Grant GB-17614. Leonard R. Hersh, Anita J. Gavazzi, Marion Nydam, Howard M. Schanker, Hetty Rosenzweig, Leonard Rosenzweig, and Jessica Utts assisted in the collection of the data. Appreciation is due Daniel L. Yeager for his contribution to Experiment III and for a critical reading of the manuscript.

[2] Requests for reprints should be sent to Jack Richardson, Department of Psychology, State University of New York, Binghamton, New York 13901.

EXPERIMENTS I AND II

Method

Lists. The compound list consisted of six trigrams as stimuli and the digits 1 through 6 as the responses. The six trigrams (CCCs) were constructed from 18 consonants so there was no overlap of letters, and it was possible to learn the response to any of the three stimulus letters or to any combination of the three letters. A second paired-associate list, consisting of six vowels as the stimuli and six three-letter words as responses, was learned either prior to (Experiment I) or following (Experiment II) the compound-list learning in order to assess the learning ability of *S*s and to permit dividing them into fast and slow *S*s.

Subjects. The *S*s were college students and were assigned as they appeared at the laboratory to conditions which had been listed in counterbalanced orders. The *S*s in Experiment I were not necessarily naive but those in Experiment II had not previously served in a verbal-learning experiment. One *S* in Experiment I requested to leave before completing the experiment. Eleven *S*s in Experiment II failed to successfully complete the experiment, 10 because of malfunctions of the memory drum and the other because of a criterion error by *E*. These *S*s were replaced by assigning the next *S* to appear at the laboratory to that condition.

Design. The *S*s in Experiment I learned the compound list to a criterion of four correct responses on a single trial (4/6), one perfect trial (6/6), two successive perfect trials plus three additional trials (2p + 3), or two successive perfect trials plus seven additional trials (2p + 7). Eighteen *S*s were assigned to each of the four degrees of learning.

In Experiment II the 2p + 3 criterion was omitted and 24 *S*s assigned to each of the other three degrees of learning of the compound list.

Procedure. In Experiment I all *S*s learned the practice (vowel–word) list to a criterion of two successive perfect trials and then learned the compound (CCC–digit) list to a criterion of either 4/6, 6/6, 2p + 3, or 2p + 7. Both lists were presented on a Stowe memory drum by the study–test method at a 2-sec. rate with a 2-sec. intercycle interval.

Following compound-list learning, the 18 letters from the six stimulus trigrams were presented individually for three recall trials. The position of the letters in the trigrams was counterbalanced within trials, and a different order of the 18 letters was used for each trial. An equal number of *S*s within each condition started recall with each of the three orders. Single letters were printed on cards and *S*s were required to turn the cards up one at a time and attempt to give the digit which had been paired with the letter during learning. If they did not know the correct response, they were required to guess one of the digits 1 through 6. The *S*s were given no information concerning the correctness of the responses during the three recall trials and all responses were recorded by *E*.

The procedure was the same in Experiment II, except the compound list was learned to a criterion of 4/6, 6/6, or 2p + 7, and the vowel–word list was learned following learning and recall of the compound list.

Results

Learning. In Experiment I the mean number of trials required to learn the practice list to the criterion of two successive perfect trials were 13.50, 10.61, 12.83, and 12.67 for the 4/6, 6/6, 2p + 3, and 2p + 7 conditions, respectively, and the differences were not significant, F (3, 68) < 1.00. There was no evidence that the groups of *S*s within conditions differed in learning ability. In Experiment II the same list required 8.75, 7.38, and 5.21 mean number of trials to learn to criterion for the 4/6, 6/6, and 2p + 7 conditions, respectively, and the differences were not significant, F (2, 69) = 2.79, $p > .05$. The trend was probably produced by the different degrees of learning of the compound list which preceeded the vowel–word list in Experiment II.

All *S*s learned the compound list to a criterion of 4/6 or higher, and the learning of the groups was compared on the trials required to attain this criterion. In Experiment I the mean number of trials required to learn the compound list to a 4/6 criterion were 3.50, 4.11, 3.67, and 3.78 for the 4/6, 6/6, 2p + 3, and 2p + 7 conditions, respectively, F (3, 68) < 1.00. The mean number of trials to the criterion of 4/6 in Experiment II were 4.46, 2.92, and 3.75 for the 4/6, 6/6, and 2p + 7 conditions, respectively, F (2, 69) = 2.88, $p > .05$. Again, there was no evidence that the groups of *S*s within experiments differed in learning ability.

Recall. There were three recall trials and the mean number of correct responses to the 18 letters were 7.69, 7.86, and 7.85 for Recall Trials 1, 2, and 3 of Experiment I. The corresponding means were 8.05, 8.33,

TABLE 1

Mean Number of Recall Responses Correct on Two out of Three (2/3) Recall Trials

Degree of learning	Speed of learning					
	Experiment I			Experiment II		
	Fast	Slow	Combined	Fast	Slow	Combined
4/6	5.56	5.11	5.33	4.83	4.17	4.50
6/6	7.33	6.44	6.89	7.75	5.92	6.83
2p(perfect) + 3	8.78	7.44	8.11	—	—	—
2p + 7	9.56	7.22	8.39	12.42	9.58	11.00
Combined	7.81	6.56	7.18	8.33	6.56	7.44

and 7.99 for Experiment II. These differences did not approach significance in either case so there was no evidence that the successive recall trials had any effect on the amount recalled.

The other recall analyses were based on a two out of three criterion, i.e., a recall response to a letter was counted as correct only if the response was correct on at least two of the three recall trials. This seems to be a better criterion than basing the analyses on a single recall trial (Richardson, in press), because it reduces the differential effects of guessing. The *S*s, within each condition of each experiment, were divided into equal groups of fast and slow learners on the basis of the trials to criterion on the vowel–word list. The analyses were based on two between-*S* variables in each experiment: speed of learning and degree of learning.

The mean number of recall responses correct to the 2/3 criterion are presented in Table 1. The trends are the same in both experiments. The correct recall responses increase as degree of learning increases and the fast *S*s recall more than the slow. The effect of degree of learning was confirmed by the analyses of variance, $F\ (3,\ 64) = 4.54$, $p < .01$, and $F\ (2,\ 66) = 33.01$, $p < .01$, in Experiments I and II, respectively. The tendency for fast *S*s to give more correct recall responses was significant only in Experiment II, $F(1,\ 64) = 3.65$, $p > .05$, and $F\ (1,\ 66) = 7.22$, $p < .01$, in Experiments I and II, respectively. The F for the interaction was less than one in both cases.

The mean number of different recall responses correct to the 2/3 criterion are presented in Table 2. This is the mean number of trigram stimuli, out of the possible six, for which one or more of the three letters elicited the correct response on at least two of the three recall trials. The increase in the mean number of different correct responses as degree of learning increased was significant for both experiments, $F\ (3,\ 64) = 5.95$, $p < .01$, and $F\ (2,\ 66) = 28.75$, $p < .01$, for Experiments I and II, respectively. The tendency for more different correct responses by the fast *S*s was not significant, $F\ (1,\ 64) < 1.00$, and $F\ (1,\ 66) = 1.17$, and the Fs for the interaction were less than one in both cases.

Tables 1 and 2 show that degree of learning and speed of learning have a greater effect on the mean number of correct recall responses than on the mean number of different correct recall responses. From this, it is to be expected that the efficiency of cue selection decreases as degree of learning increases and that the fast *S*s are less efficient than the slow *S*s.

Efficiency. The efficiency score is the percentage of the total correct recall responses that would be necessary for *S* to give a single correct response to each trigram stimulus which is represented by a correct recall to one or more of the three stimulus letters. The efficiency score was computed for each *S* by dividing the number of different correct recall responses by the total number of correct recall responses and multiplying the result by 100. The recall response to a single letter was scored as correct only if it was given correctly on

TABLE 2

Mean Number of Different Recall Responses Correct on Two out of Three (2/3) Recall Trials

Degree of learning	Speed of learning					
	Experiment I			Experiment II		
	Fast	Slow	Combined	Fast	Slow	Combined
4/6	4.22	3.89	4.06	3.50	3.25	3.38
6/6	4.56	4.89	4.72	4.50	4.25	4.38
2p(perfect) + 3	5.67	5.11	5.39	—	—	—
2p + 7	5.56	5.33	5.44	5.92	5.58	5.75
Combined	5.00	4.81	4.90	4.64	4.36	4.50

two or more of the three recall trials. The mean efficiency scores are presented in Table 3 for both experiments. The higher efficiency for the slow *S*s was not significant in Experiment I, F (1, 64) = 3.46, but was significant at the .05 level in Experiment II, F (1, 66) = 6.23. Correspondingly, the decreased efficiency with increased degree of learning was not significant in Experiment I, F (3, 64) = 1.55, but was significant at the .01 level in Experiment II, F (2, 66) = 9.98. The *F*s for the interactions were both less than one. The major difference in the efficiency scores from the two experiments was the lower efficiency in Experiment II for the 2p + 7 criterion of learning.

The six pairs in the compound list were separated into three easy and three difficult pairs for each S on the basis of the number of correct responses given to the trigram during the learning trials. When necessary, ties were assigned as easy or difficult with the aid of a table of random numbers. For each S, an efficiency score was computed for the easy pairs and for the difficult pairs. In the groups which learned to the 4/6 criterion, many *S*s did not give the correct response on two of the three recall trials to any of the nine letters from one of the categories, and an efficiency score could not be computed for these *S*s. For this reason, the groups which learned to the 4/6 criterion were not included in the analysis of the efficiency of easy and difficult items. In addition, and for the same reason, the efficiency scores could not be computed for three *S*s, one in Experiment I and two in Experiment II, in the groups which learned to the 6/6 criterion. The mean efficiency scores for the conditions were substituted for these scores in computing the analyses of variance.

TABLE 3

Mean Efficiency Scores Based on Two out of Three (2/3) Correct Recall

Degree of learning	Speed of learning					
	Experiment I			Experiment II		
	Fast	Slow	Combined	Fast	Slow	Combined
4/6	81.8	81.6	81.7	76.2	83.1	79.7
6/6	66.1	78.7	72.4	64.6	77.0	70.8
2p(perfect) + 3	67.9	74.6	71.2	—	—	—
2p + 7	65.8	76.9	71.3	51.5	63.1	57.3
Combined	70.4	77.9	74.2	64.1	74.4	69.2

The mean efficiency scores for the easy and difficult items in Experiment I were 70.8 and 78.8, respectively, and the difference was significant, F (1, 48) = 5.68, $p < .05$. The corresponding means for the easy and difficult items in Experiment II were 62.1 and 68.5, but the difference was not significant, F (1, 44) = 3.16. The interaction of speed of learning, degree of learning, and difficulty of items was also significant in Experiment I, F (2, 48) = 3.70, $p < .05$.

EXPERIMENT III

The trends in the two previous experiments were consistent in that slow learners were more efficient than fast learners and difficult items produced higher efficiency than easy items. The efficiency was high, about 80, at the 4/6 criterion of learning the compound list and decreased to about 70 at 6/6. There was no evidence of a further decrease in efficiency with overlearning in Experiment I, but there was decreased efficiency with overlearning in Experiment II.

The discrepancy in the effect of overlearning and the lack of statistical significance in some cases demanded further investigation of the variables. In addition, the somewhat higher efficiency in Experiment I suggested that learning a vowel–word list prior to learning the compound list might increase the efficiency. Experiment III extended the amount of overlearning and varied the type of stimuli in a paired-associate list learned prior to the compound-list learning.

Method

Design. The design was a 3 × 4 factorial with 24 *S*s in each of the 12 groups. The compound list was the same as in the two previous experiments and was learned to a criterion of two successive perfect trials (2p), two successive perfect trials plus seven additional trials (2p + 7), or two successive perfect trials plus fourteen additional trials (2p + 14). A control group learned only the compound list and each of the other three groups learned one of three paired-associate lists prior to learning the compound list. All three prior lists consisted of six pairs and the responses were the same three-letter words in each list. The three lists differed in the types of stimuli. One list had single vowels as stimuli and was the same list used in the two previous experiments. A second list had single four-letter words as stimuli and a third had three four-letter words for each stimulus. This variation was an attempt to determine if single-component and compound stimuli had differential effects on selection in the second list.

The *S*s were college students and had not previously served in a verbal-learning experiment. The conditions were listed in counterbalanced orders and *S*s assigned in sequence as they appeared at the laboratory. In addition to the 288 *S*s who completed the experiment, 15 *S*s were dropped; 4 due to malfunction of the memory drum and 11 due to errors by *E* or *S*.

Procedure. The procedure was the same as in Experiment I. The lists were presented by the study–test method on a Stowe memory drum at a 2-sec. rate with a 2-sec. intercycle interval. The prior list was learned to a criterion of two successive perfect trials. The compound list was learned to the specified criterion and was followed by the three recall trials to the single-letter stimuli.

Results

Learning. The first list with vowels, words, and three words as stimuli required 11.81, 8.29, and 14.46 mean number of trials, respectively, to learn to the criterion of two successive perfect trials, $F\ (2,\ 207) = 14.19$, $p < .01$. As usual, the list with compound stimuli was somewhat more difficult than the lists with single components as stimuli. There was also a significant interaction between type of prior list and degree of learning but, since degree of learning was a dummy variable at this point, the interaction must be attributed to sampling error.

TABLE 4

Mean Number of Correct Responses on Each Recall Trial for Experiment III

Degree of learning	Trial 1	Trial 2	Trial 3	Trial Combined
2p	9.11	8.93	9.04	9.03
2p(perfect) + 7	9.59	9.92	10.42	9.98
2p + 14	10.76	10.58	11.16	10.83
Combined	9.82	9.81	10.20	9.95

The compound list was learned to at least a criterion of two successive perfect trials by all *S*s and an analysis of variance was computed on the trials required to attain this criterion. The only significant source of variance was the type of prior list, $F\ (3,\ 276) = 10.11$, $p < .01$. The mean number of trials were 9.15, 6.15, 5.75, and 5.38 for the groups with no list, the vowel stimuli, the single-word stimuli, and the three-word stimuli, respectively. The major difference was due to slower learning by *S*s who learned only the compound list and this simply shows a practice effect due to prior paired-associate learning.

Recall. The analysis of the number of correct responses on the recall trials showed that degree of learning, $F\ (2,\ 276) = 10.79$, $p < .01$; recall trials, $F\ (2,\ 552) = 6.69$, $p < .01$; and the Recall Trials × Degree of Learning interaction, $F\ (4,\ 552) = 2.55$, $p < .05$, were significant sources of variance. The mean number of correct responses are presented in Table 4 for each recall trial and each degree of learning. The mean number of correct responses increased as degree of learning increased, and there were somewhat more correct responses on Trial 3 than on Trial 1 or 2.

In the rest of the analyses of recall, as in the previous experiments, a response is considered correct only if it was correct on at least two of the three recall trials. The type of prior list was not a significant source of variance in any of the recall analyses. Thus, the recall results are presented only for the nine conditions with learning prior to the learning of the compound list and *S*s were divided into fast and slow on the basis of the trials to criterion on the first list. In no case did the omission of the three groups without prior learning change the means appreciably.

The mean number of recall responses correct to the criterion of two of the three recall trials are shown in Table 5. The mean number correct increased as degree of learning increased, $F\ (2,\ 198) = 7.01$, $p < .01$, and fast *S*s recalled more than slow *S*s, $F\ (1,\ 198) = 17.26$, $p < .01$. No other sources of variance approached significance.

The mean number of different responses correct to the criterion of two of the three recall trials are also presented in Table 5. The trends were the same as for the total correct recall responses but the differences were much smaller. Degree of learning and speed of learning were both significant beyond the .01 level, and no other source of variance approached significance. However, a large proportion of *S*s in the 2p + 14 degree of learning conditions had perfect recall scores of six, and this reduced the variance in these groups so that the analysis of variance on the number of different correct recall responses should not be taken too seriously.

As in the previous two experiments, comparison of the mean number of correct recall responses and mean number of different correct recall responses suggested that efficiency decreased as degree of learning increased and that the slow *S*s were more efficient than the fast *S*s.

Efficiency. As in the previous experiments, the recall to a stimulus letter was scored as correct only if it was correct on two or more of the three recall trials, and efficiency scores were computed on the basis of this criterion of recall. As before, the efficiency scores were computed for each *S* by dividing the number of different correct recall responses by the total number of correct recall responses and multiplying the result by 100.

The first analysis of efficiency included all groups, and the type of list learned prior to the compound-list learning was not a significant source of variance. The mean efficiency scores were 58.6, 62.9, 64.9, and 64.8 for the conditions with no prior learning, vowels as stimuli, single words as stimuli, and compounds of three words as stimuli, respectively. The condition with no prior learning had a lower mean efficiency than the conditions with prior learning (58.6 vs. 64.2) and this comparison was significant, $F\ (1,\ 276) = 5.74$, $p < .05$. This indicates that learning a paired-associate list increases the efficiency of subsequent selection and that the increase does not depend upon the type of stimuli in the paired-associate list.

TABLE 5

MEAN NUMBER OF RECALL RESPONSES AND MEAN NUMBER OF DIFFERENT RECALL RESPONSES CORRECT FOR EXPERIMENT III

Degree of learning	Speed of learning					
	Recall responses			Different recall responses		
	Fast	Slow	Combined	Fast	Slow	Combined
2p	9.33	7.53	8.43	5.53	5.25	5.39
2p(perfect) + 7	10.25	8.50	9.38	5.61	5.17	5.39
2p + 14	11.30	9.50	10.40	5.89	5.67	5.78
Combined	10.30	8.51	9.40	5.68	5.36	5.52

TABLE 6

MEAN EFFICIENCY SCORES FOR EXPERIMENT III

Degree of learning	Fast *S*s			Slow *S*s		
	Easy	Difficult	Combined	Easy	Difficult	Combined
2p(perfect)	62.9	68.4	65.7	76.9	75.0	76.0
2p + 7	61.1	62.4	61.7	62.4	76.0	69.2
2p + 14	58.7	60.3	59.5	63.4	67.4	65.4
Combined	60.9	63.7	62.3	67.6	72.8	70.2

In the other analysis of efficiency scores, *S*s from the conditions with no learning prior to the learning of the compound list were discarded. The *S*s within each of the remaining nine conditions were divided into equal numbers of fast and slow *S*s based on the trials required to learn the first list to criterion. For each *S* the six pairs in the compound list were divided into three easy and three difficult items based on the number of correct responses during the learning trials. The result was three between-*S* variables (three types of prior list, three degrees of learning the compound list, and two speeds of learning the prior list) and one within-*S* variable (two levels of difficulty of items). Efficiency scores could not be computed for 2 of the 216 *S*s because of lack of recall to the 2/3 criterion, and the mean score for the condition was substituted for these scores in the analysis.

The type of prior list and interactions did not approach significance so these were combined, and the mean efficiency scores for each degree of learning, speed of learning, and item difficulty are presented in

Table 6. Efficiency decreased as degree of learning increased, $F\ (2,\ 198) = 4.01$, $p < .05$; slow *S*s were more efficient than fast *S*s, $F\ (1, 198) = 10.41$, $p < .01$; and difficult items produced higher efficiency than easy items, $F\ (1, 198) = 5.90$, $p < .05$. None of the interactions approached significance. These results confirmed the trends in the previous experiments.

GENERAL DISCUSSION

Although the effects were small, the paired-associate learning ability of *S*s was inversely related to the efficiency of cue selection and the difficulty of items directly related to efficiency in all three experiments. There was a small increase in the efficiency of cue selection in Experiment III due to prior learning of a paired-associate list. However, the increased efficiency did not depend upon whether the stimuli in the prior list were single components or compounds. Apparently, practiced *S*s are slightly more efficient without regard to the type of practice stimuli.

The efficiency of cue selection decreased as degree of learning increased. Considering all three experiments (Tables 3 and 6), the mean efficiency was about 80 at the 4/6 criterion, decreased to about 71 at 6/6, and then decreased to around 62 at the 2p + 14 criterion. This decrease in efficiency is, however, an empirical result and is not necessarily a basis for concluding that a theoretical selector mechanism becomes less efficient, or relaxes, as degree of learning increases.

The recall to the components reflects the result of the learning trials and it is assumed that learning of component–response pairs is the result of the attention given to those pairs. It is also assumed that attention may vary in the sharpness of the focus on a component and that the object of the focus may change. Thus, unnecessary component–response pairs may be learned because the focus on one component does not entirely exclude other components or because the focus is shifted to a different component after learning to one component of the compound.

First, consider the decrease in efficiency from the 4/6 to the 6/6 criterion of learning. Although the same analysis can be applied to both Experiments I and II with similar results, the data from Experiment I will be used as the example. It required 3.50 and 6.06 mean number of learning trials in Experiment I to attain the criteria of 4/6 and 6/6, respectively. As shown in Table 3, this mean difference of 2.56 learning trials produced a decrease of 9.3 in the efficiency score, from 81.7 to 72.4. Although this decrease in efficiency shows that a larger proportion of the learning from the 4/6 to the 6/6 criterion was unnecessary component–response pairs than was the case prior to the 4/6 criterion, the efficiency during the trials between criteria cannot be estimated by comparing the two efficiency scores. For this, the combined means in Tables 1 and 2 will be used. The mean number of correct recall responses (Table 1) increased from 5.33 to 6.89 from the 4/6 to the 6/6 criterion of learning, and the mean number of different correct recall responses (Table 2) increased from 4.06 to 4.72. The increase in the number of different correct recall responses (.66) divided by the increase in the number of correct recall responses (1.56) shows that 42% of the component–response pairs were necessary to produce a corresponding increase in recall to the trigrams. The other 58% of the pairs acquired during the 2.56 trials required to learn from the 4/6 to the 6/6 criterion were not necessary. The same computations at the 4/6 criterion show that 76% (4.06/5.33) of the component–response pairs learned were necessary and 24% were not necessary. The increase from 24% to 58% of the learning which was not necessary indicates that attention has changed in some fashion during the trials from the 4/6 to the 6/6 criterion.

Second, consider the decrease in efficiency with degree of overlearning in Experiment III shown in Table 6. Combining fast and slow *S*s as well as easy and difficult items gives the efficiency scores of 70.8, 65.5, and 62.4 for the 2p, 2p + 7, and 2p + 14 criteria, respectively. Examination of Table 5 shows that the decrease in efficiency is produced by the fact that the mean number of different correct recall responses is approaching the maximum of six with very little increase over the 14 overlearning trials, while the mean number of correct recall responses continues to increase about one in every seven overlearning trials. This suggests that attention continues to operate much the same throughout overlearning and that the decrease in efficiency is due to the performance limitation on the number of different responses recalled.

The data from these three experiments are compatible with the view that the sharpness of the focus of attention does not change with the various stages of learning. Early in learning, *S*s focus on single component–response pairs but without any consistent rule of selection. During this period fast *S*s learn more than slow *S*s, and by definition, more easy items are learned than difficult. After a few trials, *S*s apply a consistent rule of selection, e.g., select the first letters in the trigrams, and proceed to learn the responses to those components. Since learning has already occurred to other components and this learning is retained, the learning according to the rule decreases the efficiency and decreases it more for fast *S*s and for easy items. After the rule of selection is applied, the focus of attention is on the selected components and does not change throughout overlearning. If this interpretation of the data is correct, inducing *S*s to select according to a rule from the very first of learning should eliminate the difference in efficiency between fast and slow *S*s and between the easy and difficult items. It should also eliminate the sharp drop in efficiency prior to complete learning of the compound list.

Within the framework presented above, it is difficult to compare the results of overlearning with those from previous studies. It is clear that overlearning of the compound list produces an increase in the number of effective cues. However, it must be shown that the rate of learning of unnecessary component–response pairs increased during overlearning in order to be evidence for any relaxation of, or change in, the selector mechanism. The presentation and analysis of results in most studies do not make the necessary comparisons.

Apparently, only James and Greeno (1967) have presented clear evidence for relaxation of the selector mechanism with overlearning. In their study there was no increase in the cue effectiveness of the trigrams from the 4/8 to 8/8 learning criterion but there was an increase with overlearning. The present study does not settle this question but it does present a method of analysis which may help obtain relevant evidence.

REFERENCES

BERRY, F. M., JOUBERT, C. E., & BAUMEISTER., A. A. Single-letter cue selection and degree of paired-associate learning in retardates. *Journal of Experimental Psychology*, 1971, **88**, 196–204.

HOUSTON, J. P. Stimulus selection as influenced by degrees of learning, attention, prior associations, and experience with the stimulus components. *Journal of Experimental Psychology*, 1967, **73**, 509–516.

JAMES, C.T., & GREENO,. J. G. Stimulus selection at different stages of paired-associate learning. *Journal of Experimental Psychology*, 1967, **74**, 75–83.

LOVELACE, E. A., & BLASS, E. M. Utilization of stimulus elements in paired-associate learning. *Journal of Experimental Psychology*, 1968, **76**, 596–600.

MARTIN, E. Stimulus component independence. *Journal of Verbal Learning and Verbal Behavior*, 1971, **10**, 715–721.

RICHARDSON, J. Cue effectiveness and abstraction in paired-associate learning. *Psychological Bulletin*, 1971, **75**, 73–91.

RICHARDSON, J. Stimulus selection in associative learning. In C. P. Duncan, L. Sechrest, & A. W. Melton (Eds.), *Human memory: Festschrift for Benton J. Underwood*. New York: Appleton-Century-Crofts, in press.

RICHARDSON, J., & STANTON, S. Some effects of learning to a set of components on stimulus selection. *American Journal of Psychology*, in press.

SIEGEL, A. W., & STEVENSON, H. W. Incidental learning: A developmental study. *Child Development*, 1966, **37**, 811–817.

WICHAWUT, C., & MARTIN, E. Selective stimulus encoding and overlearning in paired-associate learning. *Journal of Experimental Psychology*, 1970, **85**, 383–388.

(Received July 24, 1972)

Journal of Experimental Psychology
1973, Vol. 98, No. 2, 404–412

DIFFERENCES IN CODING PROCESSES RESPONSIBLE FOR INTERFERENCE IN PAIRED-ASSOCIATE LEARNING[1]

WILLARD N. RUNQUIST[2]

University of Alberta, Edmonton, Alberta, Canada

Paired-associate lists in which consonant trigram stimuli shared two letters were learned by anticipation at either a 1:2 or 3:2 sec. rate. The letters making up the stimuli were arranged so that both distinctive and common letters were in identical positions in all items, only distinctive letters were in the same position, or both kinds of letters were in different positions. Half of the *S*s in each list were instructed as to the structure of the stimuli. These instructions enhanced performance only if the distinctive letter was in the same position in all items. Several analyses and a second experiment suggested that this deficit resulted from differences in stimulus coding time during both learning and recall.

When the human learner is required to associate each of several different events with a different cue for recall, the problem consists of not only learning the correct associations, but of alleviating the interference produced by the concomitant presence of other associations learned in the same context.

It has recently been argued (Runquist, 1970, in press) that the increased interference which occurs when cue stimuli are formally similar is resolved by establishing and utilizing discriminative cue codes and not by the direct weakening of interfering associations. Thus, any condition which delays or increases the difficulty of establishing stable discriminative codes for the cue stimuli should allow formal similarity to exert powerful inhibitory effects on learning. Conversely, conditions which facilitate the acquisition of such codes should reduce the effects of similarity.

Furthermore, the evidence has indicated that, at least in some cases, the coding process appears to function like a rule for generating distinctive codes and not merely like a mediating response associated with each specific stimulus (Runquist, in press). Thus, when fewer rules are necessary to generate distinctive codes or when the rules are more salient, interference should be less persistent.

Using unintegrated consonant trigrams as stimuli in paired-associate lists, Runquist (1970) has shown this to be the case. When stimuli shared two letters and these letters were in the same position in each stimulus, learning was as rapid as when there were no shared letters at all. Scrambling the position of the redundant letters resulted in considerable increase in difficulty even though the discriminative letters held the same position in all stimuli. When the position of the discriminative letter also differed among the items, interference was also substantial.

It should be noted that, in each of the above cases, a single verbalizable rule could be used to define the discriminative letter. In the first two cases, the code letter was in the same position in all items. In the third case, the code letter could be obtained by selecting the letter which was not one of the two redundant letters. It appears, then, that inconsistency in the positions of either the redundant letters or the code letters results in difficulty in discovering a rule which will generate a truly distinctive code for each cue stimulus in the set.

Of course in the usual verbal learning experiment, conditions make it exceedingly difficult for the learner to formulate such a rule due to the fact that the individual items are presented singly and at a rapid rate. Suppose, however, *S* were taught the structure of the stimulus set (i.e., how to specify the code letter) prior to learning. This ought to eliminate any differences due to difficulty in discovering the rule, particularly if sufficient time is allowed on study trials to encode items appropriately.

---

[1] This research was supported in part by Grant APA-88 from the National Research Council of Canada. Marva Blackmore supervised the data conditions. The *E*s were Heike Jurgens, Marc Prefontaine, Adele James, Sherry Banks, John Watt, John Schmidt, David Miller, and Robin Smith.

[2] Requests for reprints should be sent to Willard N. Runquist, Department of Psychology, University of Alberta, Edmonton, Alberta, Canada.

In the experiments reported here, *S*s were shown examples of stimulus sets having redundant and code letters arranged as they would be in the lists they were about to learn and were given a detailed description of the structure of that list. If the effects of structural complexity described above are due to differences in the difficulty of discovering a sufficient coding rule, the effects ought to be eliminated under instructed conditions.

Ample study time is not the only consideration, however. Typically, experiments involve restrictions on retrieval time which may make it difficult for *S* to activate the rule, generate the appropriate discriminative code, and still have sufficient time for retrieval and recall. This would be particularly important should code generation take longer with some rules than others. For this reason, the first experiment also involved variation in the time allotted for recall.

EXPERIMENT I

Method

Subjects and design. The *S*s were 240 introductory psychology students who volunteered for this experiment to obtain course credit. Each *S* learned a single paired-associate list by the anticipation procedure. The design involved four variables in a factorial combination. There were three similarity conditions, two anticipation times, and two instructional conditions replicated with two sets of lists. The experiment was conducted as two consecutive studies, the first with a short anticipation time (1 sec/item) and the second with a long anticipation time (3 sec/item). Within each experiment, *S*s were assigned to conditions as they appeared at the laboratory according to a prearranged scheme which randomized the order of conditions within blocks containing each condition once. Data from an additional 29 *S*s were collected but not used. Reasons for deleting these *S*s varied, but several failed to reach a performance criterion in the 3-sec. condition. They were, however, more or less equally distributed among the various conditions at this rate. All *S*s were tested by carefully trained undergraduate *E*s.

Materials. The lists consisted of six pairs in which the stimuli were consonant trigrams of low association value and the responses common one-syllable nouns. In all lists, the trigrams were constructed from eight different consonants with two consonants occurring in all items. Three different structures for the trigram set defined the variation in similarity. In one set of stimuli (CR), both the code letter and the redundant letters were in the same position in all items with the code letter always being in the middle position (e.g., JMV, JBV, JSV, JPV, JTV, JXV). In the second set (C), the code letter was always in the middle position, but the redundant letters were not consistently placed from item to item (e.g., JMV, VBJ, VSJ, JPV, JTV, VXJ). In the third set (R), the position of the code letter was inconsistent from item to item. This, of course, necessarily produced differences in the positions of redundant letters, but the order of these letters was always the same (e.g., BJV, SJV, JMV, JTV, JVP, JVX). The replication lists had the same structure as the above with the code letters being H, D, L, W, C, R, and the redundant letters Z and Q. Two different sets of response terms were also used, one with each replication. Half of the *S*s in each group used each replication list.

Procedure. Materials were presented on two adjacent IEE 80000 display cells located at approximately eye level in a booth. Stimuli were programmed by punched paper tape and were timed by a motor-driven cam in conjunction with a Hunter interval timer. Learning was by anticipation at either a 1:2 or 3:2 sec. rate with no intertrial interval and with five different orders. Prior to learning the experimental list, all *S*s learned a practice list consisting of nonsense shapes and two-digit numerals for 10 trials. This list was presented at the same rate as that used in learning the experimental list.

After learning the practice list, those *S*s who were not instructed (NI) began work on the experimental list after a brief instruction that they would now learn pairs of letter combinations and words. The instructed (I) groups were first shown a card on which was printed a set of consonant trigrams structured in the exact same manner as the experimental set to which they had been assigned but not containing any letters from this set, and *E* carefully pointed out how this list was structured, i.e., verbalized the simple rule for list construction. The example list was always the list used in the other replication condition. After it was clear that *S* understood the structure, he was informed that he would now learn a list just like the one he had seen but consisting of different letters. In the case of Lists CR and C, he was told that the code letter would always be in the middle position between J and V (Z and Q), with the additional note to those *S*s learning List C that the positions of J and V would not be the same in all items. In the case of List R, *S*s were told that the stimuli would consist of J and V (Z and Q) and another letter whose position would be different in the different items.

All *S*s then received five anticipation trials. Following Trial 5, an *S*-paced recognition–recall test was given in which the six list stimuli and 24 distractor trigrams were presented to *S* as a deck of cards, each card containing a single trigram. The *S* turned over the cards one at a time. When he came to a trigram which he thought was in the list, he called it out and tried to recall the response term associated with it. Distractors were chosen so that six of them had the redundant letters interchanged in position, six had the code letter changed in position, six had one of the redundant letters changed, and six had the code letter changed.

Following a single run through the deck, all *Ss* resumed anticipation learning. The 1-sec. group was given 15 more trials or 4 perfect trials, whichever came first. The 3-sec. group was run until they had reached eight consecutive perfect trials, regardless of how long it took, and then were given 10 trials at a 1:2 sec. rate. The purpose of this phase was to determine whether complete mastery of the list permanently reduced interference or whether interference was still present and would become manifest under greater time pressure.

Results

Several methodological issues may be dispensed with before dealing with the main findings. The first issue is the comparability of samples of *Ss* in various conditions. Separate analyses of variance were conducted on the 1-sec. and 3-sec. conditions. Within each condition, differences among groups were not significant, F (5, 114) < 1.00, in both cases. Because of the different anticipation times, however, a comparison between these two sets of conditions is not meaningful, so that their comparability cannot be assured. Since the data were collected at a 1-sec. rate prior to those at the 3-sec. rate, and since many of the *Ss* who were discarded were those who failed to reach the stringent criterion in the 3-sec. group, sample differences between the two rate conditions are possible. Such differences are unlikely, however. First, the 12 *Ss* who failed to reach criterion in the 3-sec. groups differed only slightly from the other *Ss* in mean total correct on the practice list (34.6 vs. 37.8 correct responses). Second, a comparison of the first and last 24 *Ss* to be run within each of the 1-sec. and 3-sec. conditions did not produce substantial differences (21.6 vs. 22.0 correct for 1-sec. *Ss*, and 35.9 vs. 40.0 for 3-sec. *Ss*). Finally, an analysis of both experiments with *Ss* stratified by practice list performance did not result in essentially different trends at any level. Consequently, any possible population differences would appear to affect only the main effect of test time which is theoretically trivial.

A second issue deals with the fact that the R list contains items with the code letter in all three possible positions, while the CR and C groups always have the code in the middle position. Since previous research (Runquist, 1971) has shown that code-letter position has significant effects on performance, differences among similarity conditions might reflect these effects. This argument may be allayed by comparing the two items from the R list which had code letters in the middle position with the same two items in the CR and C conditions. No differences in percentage correct for this subset and the total set were apparent under any condition. Consequently, all analyses were performed on results from all items.

Acquisition. Table 1 shows the mean total correct summed over 20 trials, the mean trials to first correct response, the mean overt errors, and the ratio of overt errors to opportunities for the different experimental conditions. If *S* reached criterion prior to 20 trials, it was assumed that further trials would result in perfect performance. Although the 3-sec. groups were run to eight correct trials in a row,

TABLE 1

Acquisition Performance Measures

Group	Mean total correct	Mean trials to first correct	Mean overt errors	Error ratio
		1 sec.		
CR-NI	60.9	34.9	9.3	.16
C-NI	40.8	49.6	16.0	.20
R-NI	47.3	47.2	12.4	.17
CR-I	80.7	22.0	6.6	.07
C-I	66.4	27.8	10.1	.19
R-I	52.8	42.0	13.2	.20
		3 sec.		
CR-NI	98.2	14.9	4.9	.23
C-NI	85.0	17.2	10.5	.30
R-NI	92.2	16.3	8.1	.29
CR-I	105.5	8.6	2.9	.20
C-I	103.5	10.0	3.5	.21
R-I	95.6	13.4	6.3	.27
MS_e	321.9	206.4	92.8	

for this analysis they were scored as if the criterion were four perfect trials, but with extremely rare exceptions, no errors occurred after four perfect trials anyway.

The total correct response measure is also influenced by a disruption in performance resulting from the insertion of the recognition test after Trial 5. While generally speaking the 1-sec. groups were more disrupted than the 3-sec. groups, there did not appear to be any systematic variations in loss within each condition.

Major trends for the data in Table 1 may best be delineated with reference to significant sources of variance in the analysis. For total correct responses, the main effects of instructions, $F\ (1, 216) = 31.74$, and anticipation time, $F\ (1, 216) = 272.69$, were both significant (both $ps < .01$). Of more interest, however, are the effects of similarity, $F\ (2, 216) = 15.44$, $p < .01$, and the interaction of similarity with instructions, $F\ (2, 216) = 4.34$, $p < .025$. Under noninstructed conditions, the pattern of results was similar to that found in a previous study (Runquist, 1970) with the CR condition producing better performance than C and R conditions which were about equal. Instructions as to list structure, however, differentially facilitated these conditions. Most notable is the fact that the R condition was scarcely affected by the instructions, while both CR and C groups showed considerable facilitation. It should be noted that the C group was still somewhat inferior to the CR group, even when the critical nature of the middle letter was pointed out. While this difference largely disappeared when anticipation time was longer, the interaction of similarity with instructions and time was not significant ($F < 1.00$). No other factors were significant, although both similarity and instructions tended to show smaller effects with the 3-sec. rate. The interactions for similarity and instructions, respectively, were $F\ (2, 216) = 2.67$ and $F\ (1, 216) = 2.81$, both $ps < .10$.

TABLE 2

Recognition Test Results

Group	Mean correct recognitions	Total false recognitions	Transpose redundant letter	Transpose code letter	Change redundant letter	Change code letter
			1 sec.			
CR-NI	5.15	34	11	9	2	5
C-NI	3.05	76	44	15	7	10
R-NI	3.45	84	43	31	1	8
CR-I	3.60	85	43	5	29	8
C-I	3.20	107	72	8	18	9
R-I	3.25	76	25	31	13	7
			3 sec.			
CR-NI	5.35	29	6	9	7	7
C-NI	3.35	63	46	9	7	1
R-NI	4.10	48	22	16	6	4
CR-I	4.70	37	22	10	4	1
C-I	3.95	101	74	3	22	2
R-I	4.55	111	65	32	8	6

The mean number of trials to the first correct response provides a measure of early acquisition. In this case, the results are almost redundant with the total correct response measure. Instructions, $F\ (1, 216) = 25.10$; anticipation time, $F\ (1, 216) = 163.64$; and similarity, $F\ (2, 216) = 9.44$, were all significant ($p < .05$). In addition, the interactions of similarity and instructions with time, which were not significant with total correct responses, did achieve significance with this measure, $F\ (2, 216) = 4.20$, $p < .025$, and $F\ (1, 216) = 4.60$, $p < .025$, respectively. The interaction of similarity with instructions, however, did not reach acceptable reliability levels, $F\ (2, 216) = 2.78$, $p < .10$, but exhibited the same trend as with correct responses.

The overt error results were as unenlightening as usual. Only the main effects of similarity, $F\ (2, 216) = 4.77$, $p < .025$; instructions, $F\ (1, 216) = 6.18$, $p < .01$; and anticipation time, $F\ (1, 216) = 17.99$, $p < .01$, were significant. Essentially, CR groups made fewer errors than C or R. The error rate, however, did not differ substantially among the groups, although it was somewhat higher with 3-sec. anticipation time.

Recognition test. Several measures were available from the recognition test given after Trial 5. These results are summarized in Table 2. The first column reports mean correct recognitions (maximum = 6). The differences among the similarity conditions were significant, $F\ (2, 216) = 15.98$, $p <$

.01, and showed the usual pattern of superiority of the CR group over both C and R groups, which did not differ. Anticipation time also had its usual facilitating main effect, F (1, 216) = 9.30, $p < .01$, although in this case it is more interesting, since on the recognition test there are no time restrictions. Thus, the effect must represent some permanent residual effects of longer anticipation time on the trials preceding this test and could be due to more exposure to stimuli in the 3-sec. conditions. While the main effect of instructions was not significant ($F < 1.00$), the interaction of instructions with similarity did reach an acceptable reliability level, F (2, 216) = 5.84, $p < .01$. The effect, however, differed from that found on acquisition performance, in that only the CR group performance changed with instructions, and the effect was to *decrease* the number of correct stimulus identifications.

The total number of false recognitions made by *S*s in each group is shown in the second column of Table 2. The main effects of similarity, F (2, 216) = 12.44, and instructions, F (1, 216) = 18.20, were both significant (both ps $< .01$). The similarity effect reflected the usual pattern, while the effect of instructions was to increase false recognitions from an average of 2.7 per *S* to 4.3 per *S*. Further assessment of the false recognition data was hampered by a significant fourth order interaction among similarity, instructions, anticipation time, and lists, F (2, 216) = 4.79, $p < .05$, but a plot of this interaction revealed that the essential aspect of the significant interaction among similarity, instructions, and time, F (2, 216) = 5.57, $p < .01$, remained interpretable. In general, this reflects two trends: At 1 sec., instructions increased false recognitions for Conditions CR and C, but not for Condition R, and at 3 sec., instructions increased errors for Conditions C and R, but not for Condition CR.

The next four columns present the distribution of false recognitions of various types. First, it is clearly apparent that *S* rarely accepts a distractor in which the code letter has been changed. Moreover, under noninstructed conditions, he rarely accepts a distractor in which one of the redundant letters has been changed. Most of the false recognitions consist of items in which the position of either the code letter or the redundant letters is rearranged, with the latter predominating. The effect of instructions and its significant interaction with similarity and anticipation time are reflected in increases in various kinds of errors. At 1-sec. anticipation time, the increase in false recognitions for both CR and C results from an increase in accepting distractors in which the redundant items have been either transposed or changed. For Condition R, there is a small drop in the frequency of false acceptances for items with redundant letters transposed, but this is offset by a small increase in accepting items with redundant letters changed. Whether these latter changes are significant or not is hard to tell. They represent less than one item/*S* on the average. At 3 sec., the increase in false recognitions for instructed C and R groups, while about the same in magnitude, is somewhat different in character. Both groups show an increase in the frequency of accepting items with transposed redundant letters, but the C group also increased its acceptance rate for items with redundant letters changed, while the R group increased the frequency of errors where the code letter was changed in position. Again it should be pointed out that the effect of anticipation interval is a residual effect from the five previous learning trials in that differential temporal restrictions were not operating on the recognition test.

The *S* was also requested to attempt to recall the correct response on the recognition test. Credit for a correct recall may be given in two ways: either when *S* has correctly recognized the stimulus and recalls the appropriate response (strict), or when he has falsely accepted a distractor containing the correct code letter, but recalls the correct response to that code letter (lenient). This second method seems perfectly appropriate considering the task set for *S* in paired-associate learning (Runquist & Evans, 1972). The results on these two measures are presented in Table 3 in terms of proportion recalled (out of six possible). The recall of any given response was scored only once for each *S*, despite possible multiple false recognitions containing a particular code. He received credit for a correct recall if he recalled the correct response to any stimulus containing that code letter. For the strict recall measure, only similarity, $F\ (2, 216) = 10.59$, $p < .01$, and anticipation time, $F\ (1, 216) = 47.73$, $p < .01$, were significant. Both effects were the same as with previous measures. Neither instructions, $F\ (1, 216) = 1.40$, nor the interaction of instructions with similarity, $F\ (2, 216) = 1.66$, was significant. For lenient recall, similarity, $F\ (2, 216) = 5.83$; instructions, $F\ (1, 216) = 13.64$; and anticipation time, $F\ (1, 216) = 31.70$, were all significant (all ps $< .01$) and the interaction of similarity with instructions approached significance, $F\ (2, 216) = 2.85$, $p < .10$. The main feature of this analysis is that the essential differences among conditions hold even when recall is unpaced.

The final two columns in Table 3 present the recall data conditional on correct recognitions only (strict) or conditional on both correct recognitions and all false acceptances (lenient). These results did not differ from unconditional recall, except that after learning with the 3-sec. anticipation time conditional recall was uniformly high. Also of some interest is the fact that it made little difference whether recall was conditionalized only on correct recognitions or also included appropriate false recognitions.

TABLE 3

RECALL ON RECOGNITION TEST

Group	Proportion strict recall	Proportion lenient recall	Conditional strict recall	Conditional lenient recall
		1 sec.		
CR-NI	.54	.57	.63	.59
C-NI	.28	.42	.56	.45
R-NI	.32	.42	.57	.48
CR-I	.47	.66	.78	.71
C-I	.38	.63	.72	.70
R-I	.28	.48	.45	.49
		3 sec.		
CR-NI	.73	.80	.83	.84
C-NI	.51	.57	.85	.70
R-NI	.55	.64	.81	.76
CR-I	.73	.80	.95	.88
C-I	.66	.79	.94	.91
R-I	.67	.77	.89	.88

Reduced anticipation time. When those *S*s who had learned at a 3:2 sec. rate were switched to a 1:2 sec. rate, perfect performance was disrupted in all groups. Moreover the R groups lost more (26%) than the CR groups (16%) or C groups (18%). The difference was maintained over all 10 trials. Similarity was the only significant source of variation both on the first trial, $F\ (2, 108) = 3.96$, $p < .05$, and for total correct responses, $F\ (2, 108) = 4.19$, $p < .025$. Although these results could be due to general procedural change, it would appear to be more likely that temporal restrictions on retrieval are involved.

There is little reason to expect differential loss due to general disruption, particularly between C and R conditions where performance was equally bad prior to criterion. Moreover, the effects persisted over 10 trials.

Discussion

The effects of structural relations among stimuli are clearly demonstrated in these results. Lists in which stimulus terms shared two letters were clearly easier when the redundant letters and the code letter were in the same position in all items. When the position of either the redundant letters or both redundant and code letters varied from item to item, performance was about equally disrupted. The absence of a low-similarity group made it impossible to determine whether letter position consistency completely eliminated the effects of similarity as has been found in other studies (Runquist, 1970). While this comparison might be of some interest with the noninstructed groups, any comparison with instructed conditions would be complicated by the fact that the latter had information about the stimuli which could not be made equivalent for low-similarity lists.

Instructing *S*s as to the list structure facilitated performance as long as the code letter was in a consistent position, but had virtually no effect on correct responding in Group R when letter positions varied. Comparisons between the two groups which had consistent code-letter positions generally indicated that

Group CR, in which redundant letters also were in identical positions, continued to perform somewhat better than Group C, in which the redundant letters varied in position, although differences were often small, particularly when *S* was given sufficient time to recall. The conclusion may be made that knowledge of a rule for generating distinctive codes is insufficient to eliminate the differential effects of structure in lists with highly similar stimuli.

This conclusion depends, of course, on the assumption that *S*s did, in fact, understand the structure of the list they were to learn. Although no independent measure of this knowledge is available, it does not appear reasonable that any misunderstanding could exist even in the R condition where the rule is somewhat more complicated. Moreover, the types of false recognitions made by various groups consistently indicate that *S*s not only understood the structure, but utilized this knowledge in appropriate ways, i.e., by ignoring redundant letters in groups where position was a sufficient determiner of the code letter and ignoring position in Group R where it was not. While differences in rule understanding cannot be completely discounted, it would appear much more likely that some additional factor or factors associated with list structure affected the utilization of the rule, thus maintaining performance differences.

One possibility is that interference is involved not just in stimulus coding, but also in the associative process itself. In order to distinctively identify a stimulus in lists of the kind used here, it would appear to be necessary for *S* to have recourse to the code letter as it is the sole distinctive feature. On the other hand, the list structure does not ensure the utilization of this letter as a retrieval cue. Retrieval may consist of a sequential process in which *S* first codes the stimulus in terms of its distinctive code letter, then utilizes the code letter in conjunction with some other attribute of that simulus as a retrieval cue. Position differences, even in redundant letters, may make associative learning or retrieval more difficult.

The role of anticipation time does not appear as simple as considered earlier. Originally, it seemed obvious that utilization of coding rules would take time, and instructions might be expected to produce less effect when rather severe limitations were placed on retrieval time. In one sense, this expectation was upheld in that differences among the similarity groups in the instructed conditions were smaller at a 3-sec. than at a 1-sec. anticipation time.

It is clear, however, that differences in anticipation time do not simply affect time to retrieve. Recall performance on the unpaced recognition test yielded better performance for all groups which had a 3-sec. anticipation time during the first five acquisition trials when compared with groups having a 1-sec. anticipation time. Furthermore, while differences in recall among the similarity conditions were small after learning at a 3:2 sec. rate, Group R, in which the position of the code was inconsistent, recalled fewer items than the other two groups when they had learned at a 1:2 sec. rate. It would appear, then, that some processing of the stimuli occurs during the anticipation interval which has permanent effects, and this processing is less efficient when the code letters are not in the same position. Not all of the deficit in performance of this group can be attributed to this permanent process, however, since even after considerable overlearning at a 3:2 sec. rate, switching to a 1-sec. anticipation interval attenuated performance more when code letters were inconsistently placed.

Actually, these results are resolvable in terms of the coding-rule hypothesis. During the anticipation interval, *S* must activate the rule and generate the discriminative code as well as attempt recall. With strict time limitations, the coding process may not be completed, and thus before associative learning can proceed, *S* must complete this processing during the study time. The longer it takes to complete coding, the less "rehearsal" time is available. Longer anticipation times simply allow for more time to be devoted to associative learning during the study period.

The further implications are that generating the distinctive code takes longer when the rule is not defined by consistent position. Group R thus suffers deficiencies of two kinds: less time for retrieval and less time for associative coding, the latter becoming less serious as anticipation time is lengthened.

The critical assumption is that even knowing the rule, code generation takes more time in the case of R than CR and C. Experiment II was designed to obtain direct evidence for the assertion.

EXPERIMENT II

This experiment attempted to directly measure the latency of code generation for the three types of stimulus sets used in Experiment I. Essentially, *S* was presented with the stimuli from the list one at a time and was required to call out the code letter.

Method

Subjects and design. The *S*s were 60 introductory psychology students selected and assigned to one of three conditions in the same manner as in Experiment I. The three conditions were defined by the same three paired-associate lists used in Experiment I. Each *S* learned a single list. Immediately preceding learning, and after Trial 5 and Trial 10, the stimuli were presented singly, and *S* was asked to call out the code letter as quickly as possible. After the experiment was completed, an additional eight *S*s were run on each of Lists C and R, with the code-latency test being given only after Trial 10. All *S*s were tested by carefully trained undergraduate *E*s.

Apparatus. All items were rear projected onto a translucent screen by means of a Dunning Animatic 16-mm. filmstrip projector. Items appeared white on a dark background. For learning trials, presentation was at a 2:2 sec. rate and was controlled by a motor-driven cam. On the code-latency test, the presentation rate was unpaced. On these trials a solenoid operated diaphragm shutter was placed in front of the lens of the projector. The filmstrip was manually advanced by *E* who then exposed the stimulus by pressing a button which activated the shutter and simultaneously started a Hunter Klockounter. A Grason-Stadler voice-operated relay stopped the clock when *S* made a vocal response.

Procedure. All *S*s received 10 paired-associate learning trials on the nonsense shape–numeral list. They were then given instructions as to the structure of the stimulus sets, which were identical to those given in Experiment I. In addition, they were instructed as to the code-latency test which followed immediately. On this test, the stimuli were presented three times in three different orders. The time between presentations was somewhat irregular, being determined by how long it took *E* to record the time from the previous item and set up the next item. The *E* said "ready" out loud prior to presenting the item. No special interval was given between trials. Following this initial latency test, *S*s received five trials of anticipation learning, a second set of three code-latency trials, five more learning trials, and three final latency test trials. Five orders of presentation were used on the learning trials. There were three sets of orders for use on the three latency tests. Each *S* used each set, but the sequence was randomly determined.

Results and Discussion

Learning. There were no differences in practice list performance among the three groups ($F < 1.00$). The means for total correct responses on the 10 learning trials were 35.7, 29.8, and 26.8 for Groups CR, C, and R, respectively. Thus, the learning data replicate the findings from Experiment I at an anticipation time between those used in that experiment.

Code latency. In each group, there were 1,080 latency trials. There were 3, 6, and 32 errors made in calling out the code letter in the CR, C, and R groups, respectively. While the R group made more errors, it should be pointed out that this represents only 3% of the total test trials given, and in any single trial block of six stimuli, there was never more than one error. These data may also be offered in support of the fact that *S*s in Experiment I did understand the structure of the list, because with the exception of the instructions for responding, instruction in the two experiments was identical.

The latencies for each block of trials at each stage of practice were first averaged, thus providing each *S* with 18 measures. All analyses were performed on these means. A four-factor analysis of variance with similarity and lists as between-*S* factors and stage of practice and test trials within stages as within-*S* factors resulted in significant effects for similarity, $F\ (2,\ 54) = 51.85$; stage of practice, $F\ (2,\ 432) = 18.50$; test trials within stages, $F\ (2,\ 432) = 38.89$; and the interaction between stage of practice and test trials, $F\ (4,\ 432) = 8.81$, all $ps < .01$. There were also two interactions with lists: similarity and the similarity by practice stage interaction. Both of these simply reflect greater effects with one of the two lists and are of no consequence in interpreting the results.

The effect of major interest was the differences in latency among the three similarity groups. These means were 575, 593, and 806 msec. for CR, C, and R groups, respectively. Several analyses performed on the third test trial after 10 trials of learning dispel any doubts that this result is an artifact. It is neither a consequence of one slow item, nor of different stimuli being used in the lists. If one looks at the item with the shortest latency, Condition R maintains its disadvantage (mean latencies were 477, 469, and 604 msec. for the CR, C, and R groups, respectively). A comparison of the three different code positions in Group R produced evidence that code search in this group is serial and self-terminating in that latency increased with serial position (709, 765, and 912 msec.). Nevertheless, all three values are longer than those obtained for the CR and C items on this trial (536 and 535 msec., respectively). Without any doubt, it takes longer to obtain the code letter when its position is variable.

The other main effects are also of some interest. There was some improvement within each of the tests. The mean latencies were 716, 628, and 628 msec. This effect was largely due to slightly longer latency on the first item of the first block. Latency also showed a significant decrease with stage of practice: 697, 646, and 633 msec. after 0, 5, and 10 learning trials, respectively. The result apparently has little to do with paired-associate learning, however. First, a plot of the interaction of practice with test showed that the effect was restricted to the first test trial in the series of three. Second, two groups of *S*s ($n = 8$) who were given the code-latency test only once showed latencies after 10 learning trials almost identical to those obtained by comparable groups on the test prior to paired-associate learning.

Irrespective of these results however, the latency data are in complete accord with the hypothesis.

GENERAL DISCUSSION

Very little need be added at this point except to consider the implications of these results for a general theory of interference. Whereas the presence of rule-governed coding processes has been amply demonstrated in dealing with unintegrated verbal units in free recall (Underwood & Erlebacher, 1965; Underwood & Keppel, 1963), the fact that such mechanisms can operate to reduce interference produced by letter identity among stimuli seems now to be well established. As long as the cue codes are based on the formal characteristics of the item, at least the availability of coding rules appears to largely determine just how much interference is produced.

These results would also seem to establish the fact that being able to generate a distinctive cue code does not necessarily alleviate all interference, as performance differences were still present when such codes could be generated. The code-latency results and effects of increased anticipation time on learning and performance in the different conditions point to the critical role of processing time, and suggest that it is possible for differences in interference in both learning and retrieval to result from differences in coding time and not from differential stimulus confusion or direct associative interference.

While temporal factors can be implicated as one source of interference, they are clearly not the whole story. Differences in performance between the two groups in which code position was constant point this out most clearly. Small differences appeared throughout the learning data, which according to the code-latency results cannot be attributed to differences in time taken to generate the distinctive code. While one might speculate on what is involved, such speculation does not appear theoretically fruitful in that factors involved may be restricted to those particular structures. Suffice it to say that the false recognition data suggest rather profound differences in the knowledge about the redundant letters in the stimuli for the two groups. The critical point for a theory of interference is that even given equal availability of a distinctive code, there is impairment in learning.

The overall lesson to be learned from these experiments, however, is that various similarity conditions may produce equal amounts of interference, but the mechanisms responsible for the persistence of this interference may involve both learning and performance factors which are quite different.

REFERENCES

RUNQUIST, W. N. Structural effects of letter identity among stimuli. *Journal of Experimental Psychology*, 1970, **84**, 152–163.

RUNQUIST, W. N. Ordinal position of letter identity among stimuli. *Journal of Experimental Psychology*, 1971, **87**, 270–273.

RUNQUIST, W. N. Intralist interference and stimulus similarity. In C. P. Duncan, L. Sechrest, & A. W. Melton (Eds.), *Human memory: Festschrift for Benton J. Underwood*. New York: Appleton-Century-Crofts, in press.

RUNQUIST, W. N., & EVANS, A. Stimulus recognition and associative coding. *Journal of Experimental Psychology*, 1972, **95**, 242–244.

UNDERWOOD, B. J., & ERLEBACHER, A. Studies of coding in verbal learning. *Psychological Monographs*, 1965, **79** (13, Whole No. 606).

UNDERWOOD, B. J., & KEPPEL, G. Coding processes in verbal learning. *Journal of Verbal Learning and Verbal Behavior*, 1963, **1**, 250–257.

(Received July 24, 1972)

Journal of Experimental Psychology
1973, Vol. 98, No. 2, 413–422

STIMULUS AND RESPONSE REPETITION EFFECTS IN RETRIEVAL FROM SHORT-TERM MEMORY:

TRACE DECAY AND MEMORY SEARCH[1]

EDWARD E. SMITH[2]
Stanford University

WILLIAM G. CHASE
Carnegie-Mellon University

PETER G. SMITH
University of Wisconsin

Repetition effects were studied as a function of intertrial interval (ITI) in a four-stimulus two-response reaction time task. Repetition effects were varied by repeating (on adjacent or nonadjacent trials): (*a*) both the stimulus and the response, (*b*) only the response, or (*c*) neither. An analysis that considered repetition effects only across adjacent trials performed by Bertelson in 1965 showed the occurrence of stimulus and response repetition effects, both of which declined with ITI. Although this is consistent with the hypothesis that repetition effects are mediated by a rapidly decaying trace, a further analysis that considered repetitions over both adjacent and nonadjacent trials revealed primarily stimulus (rather than response) repetition effects, some of which did not consistently decrease with ITI. It was argued that the results of this more extensive analysis were consistent with the hypothesis that repetition effects reflect a self-terminating memory search, rather than trace decay.

In 1961, Bertelson reported a repetition effect in a choice reaction time (RT) task; responses to a stimulus were faster if that stimulus was identical to the preceding one than if it differed. This effect is important because it may mediate other RT effects found in studies of rapid retrieval from short-term memory, for example, the effect of number of alternatives (Kornblum, 1969). The purpose of this article is to attempt to determine the specific mechanism responsible for the repetition effect. But before doing so, it is first necessary to consider the locus of this effect within the sequence of processes that constitute retrieval from short-term memory (STM).

There is substantial evidence that rapid retrieval from STM is accomplished by a series of four processing stages: (*a*) The first stage deals with a preprocessing or encoding of the stimulus, (*b*) the second stage includes a categorization or comparison of the stimulus with respect to alternatives stored in STM, (*c*) the third stage deals with response selection, and (*d*) the final stage directs response execution (E. E. Smith, 1968; Sternberg, 1969). Thus, a distinction can be made between stages that process information about the stimulus (the encoding and comparison stages) and stages that are primarily concerned with selecting and executing a response. The first issue of interest is whether the repetition effect is due mainly to the former stimulus-oriented stages or to the latter response-oriented stages. Now, if the repetition effect were attributable only to response stages, then a repetition of just the response (and not the stimulus) should be sufficient to produce a repetition effect, since the operations of the response stage are being repeated. Conversely, if the repetition effect were due only to stimulus stages, then a repetition of just the response should be insufficient to produce a repetition effect. This logic suggests that one may determine, in part, the theoretical locus of the repetition effect (stimulus vs. response stages) by determining whether there is a response repetition effect, a stimulus repetition effect, or both.

---

[1] The authors are grateful to John Theios for the use of his computerized laboratory (supported by U.S. Public Health Service Research Grant MH-16711) and, more importantly, for the direction he has given to the research. The very helpful comments of Dominic Massaro are also gratefully acknowledged. This research was also partly supported by U.S. Public Health Service Grants MH-17222 and MH-19705 to the first author and MH-07722 to the second author. Part of this research was presented at the 1970 meeting of the Midwestern Psychological Association.

[2] Requests for reprints should be sent to Edward E. Smith, Department of Psychology, Stanford University, Stanford, California 94305.

An experimental separation between stimulus and response repetition effects can be accomplished by associating two stimuli with one response, and then repeating on adjacent trials either (*a*) both the stimulus and response, (*b*) only the response, or (*c*) neither. The response repetition effect may then be estimated by the RT difference between *c* and *b*, while the stimulus repetition effect is given by the difference between *b* and *a*. Bertelson (1965) was the first to use this procedure, but his results may be of dubious import in view of a confounding which permitted *S* to treat the two stimuli associated with the same response as a single stimulus (M. C. Smith, 1968). (Specifically, Bertelson, 1965, associated the stimulus digits 2 and 4 with one response and the digits 5 and 7 with a second response, thus allowing *S* to determine his response on the basis of whether the stimulus was odd or even.) More recent studies using the Bertelson procedure have avoided this problem. In these studies (Eichelman, 1970; Rabbitt, 1968), a sizable stimulus repetition effect emerges consistently; a response repetition effect may sometimes occur and may other times be replaced by an inhibitory effect, i.e., repetition of only a response leads to slower RTs than when neither the stimulus nor the response is repeated (e.g., M. C. Smith, 1968). Overall it seems that much of the Bertelson (1961) repetition effect is a stimulus repetition effect, and hence can be localized to one of the stimulus stages of retrieval.

Two explanations of the mechanism underlying the repetition effect have been offered, and in accordance with the above findings, both assume that the locus of the effect is at a stimulus stage of retrieval, namely the comparison stage. The first explanation is the "trace decay hypothesis," which holds that when a stimulus occurs *S* first compares it to the still present, but decaying, sensory trace of the preceding stimulus. If the present stimulus is identical to the preceding one and the preceding one's trace has not undergone extensive decay, then a match will occur between the trace and the stimulus. Thus *S* would not have to engage in relatively long-duration comparison processes to categorize the repeated stimulus, as he would to categorize a nonrepeated stimulus whose matching sensory trace would be unavailable (e.g., Schvaneveldt & Chase, 1969). Several investigators have tested the trace decay hypothesis by seeking a decline in the magnitude of the repetition effect as the intertrial interval (ITI) lengthens. Presumably, as the ITI lengthens, the sensory trace of the preceding stimulus decays and becomes less usable for matching the present stimulus. Some studies (e.g., M. C. Smith, 1968; Umiltá, Snyder, & Snyder, 1972) have, in fact, found that the repetition effect diminishes as the ITI lengthens. But other attempts to obtain this ITI finding have produced impressive failures (e.g., Schvaneveldt & Chase, 1969).

One possible reason for these inconsistent ITI findings is the failure of the above studies to separate the stimulus repetition effect from response effects, facilitative or inhibitory. For stimulus and response effects may be differentially influenced by ITI variations. Eichelman (1970) performed a critical experiment as he assessed the influence of ITI duration on stimulus and response repetition effects separately. In a task requiring the naming of single letter stimuli, Eichelman found that as the ITI lengthened 200–700 msec. the stimulus repetition effect decreased while the response repetition effect (a facilitative one) increased. Eichelman argued that the stimulus repetition effect is mediated by the same trace decay process that is operative in a *same–different* RT task. In the latter type of task, when a brief interstimulus interval is used, *same* RT is faster when the two stimuli are physically identical (a physical match) than when they are only nominally identical (a nominal match; Posner, Boies, Eichelman, & Taylor, 1969). Now, Posner et al. assumed that a rapidly decaying visual trace was the basis for a physical match being faster than a nominal match, and Eichelman (1970) extended the Posner et al. variation of the trace decay hypothesis to account for his stimulus repetition results. However, there is a problem with this explanation. If the stimulus repetition effect is mediated by a decaying visual trace, then when a single *different* item intervenes between the first and second occurrence of the stimulus in question, the visual trace of that stimulus may be lost (Posner et al., 1969). (Specifically, in a *same–different* task, Posner et al. found that when *S* processed extraneous information during the 500-msec. interval separating the two items in a *same* pair, physical matches were no faster than nominal matches.) Thus there should not be any sequential effect over nonadjacent trials, i.e., the Eichelman–Posner trace decay hypothesis predicts a stimulus repetition effect over adjacent trials but not over nonadjacent or remote trials. But remote repetition effects have been convincingly demonstrated; for example, Schvaneveldt and Chase (1969) showed that a response to a stimulus was faster if that stimulus was identical to the one prior to the immediately preceding stimulus.

The above review suggests the need for another explanation of repetition effects and the other explanation we shall consider will be termed the "memory search hypothesis." This hypothesis assumes that the categorization of any stimulus involves a sequential search of STM, where the order in which the memorial alternatives are compared to the stimulus is probabilistically determined by how recently these alternatives have occurred (e.g., Theios, Smith, Haviland, Traupman, & Moy, 1973). To be more concrete, consider the case where all of the alternatives' representations are maintained in STM. On any trial, after the stimulus has been encoded, STM is searched in a serial self-terminating manner for a representation that matches the stimulus, where the search order of the memorial representations is probabilistically determined by how recently the alternatives corresponding to these representations have occurred. This hypothesis clearly accounts for the adjacent stimulus repetition effect since the last alternative presented is frequently among the first ones searched. It also explains remote repetition effects since the representation of a stimulus that has occurred on Trial $n-2$ (T_{n-2}) will probably be higher in the search order on T_n than the representation of a stimulus that has not occurred in several trials (though by this line of reasoning, remote repetition effects should, on the average, be of smaller magnitude than adjacent repetition effects). Furthermore, this hypothesis explains the finding that RT continues to decrease with increasing repeated presentations of the same stimulus (e.g., Umiltá et al., 1972), as repeated presentations continually increase the probability that the representation of that stimulus is the first one checked. While an unembellished memory search hypothesis cannot account for a decrease in the stimulus repetition effects as the ITI lengthens (as Eichelman, 1970, found), the addition of one plausible assumption to this hypothesis may adequately handle this problem. It may be assumed that as the ITI lengthens, there is an increasing probability that some task-irrelevant stimulus (produced either externally or by *S* himself) will enter STM (J. Theios, personal communication, June 1972). This will lower all of the memorial representations in the search order and consequently reduce stimulus repetition effects.

Thus the trace decay hypothesis can easily account for the declining stimulus repetition effect over ITI while the memory search hypothesis handles this finding by an additional assumption; on the other hand, the memory search hypothesis can explain remote stimulus repetition effects while the trace decay hypothesis cannot. The present study attempted to clarify the status of these two hypotheses by considering both remote stimulus repetition effects and ITI effects within the same task. Presumably such a study can assess the relative validities of the two hypotheses more directly than previous experiments.[3]

METHOD

Subjects. The 22 *S*s who participated in this experiment were graduate students and junior faculty members at the University of Wisconsin. Their ages ranged 22–33 yr. and all but one of them were right handed.

Apparatus and procedure. Each *S* was run individually in a soundproof chamber and white noise was presented by earphones to mask extraneous sounds. On each trial a single digit, either 1, 2, 3, or 4, was presented. The *S*'s task was to press one button with his left forefinger if the digit was a 1 or a 2, and the other button with his right forefinger if the digit was a 3 or a 4.[4] The digits were presented via an Industrial Electronics Engineer readout while the response buttons were mounted in the arms of *S*'s chair. For each *S*, three ITIs were used: 100, 500, and 1,000 msec. The ITI was defined as the interval between the response on T_n and the onset of the stimulus on T_{n+1}. The *S* received 320 test trials in a block at each ITI, though only the last 300 trials were considered in the analyses. The order of ITIs was randomized for each *S*. Throughout the entire session the stimuli were equiprobable and were presented in random order. The usual speed–accuracy instructions ("Respond as quickly as possible without making errors.") were used, and *S* was given 76 practice trials before the 960 test trials. A Digital Equipment Corp. PDP-8 computer was programmed to control the presentation of the stimuli, control the duration of the ITI, and measure each RT to the nearest millisecond.

RESULTS

Two different analyses of mean RTs will be presented. The first considers only adjacent repetitions and is modeled after Bertelson's (1965) procedure, while the subsequent one involves an extension of Bertelson's procedure to consider remote as well as adjacent repetitions. Both analyses consider only correct responses. The overall error rate was 5% and error rates were typically positively correlated with mean RT, and hence relatively uninformative. Error frequencies will be presented with correct RT data but will not always be commented on. Also, for both analyses, tabular or graphic results are presented and any conclusion drawn from them is always supported by the outcome of an analysis of variance and often by the outcome of a subsequent statistical test.

A Bertelson-type analysis of adjacent stimulus and response repetitions. The unit in this analysis is a pair of adjacent trials and the focus is on the effect of a preceding trial, T_{n-1}, on a test trial, T_n. Accordingly, the entries in the first three rows of Table 1 present, for each ITI, mean RT and error percentage as a function of whether: (*a*) T_{n-1} and T_n involved the same stimulus and the same response (an identical [I] transition between adjacent trials), (*b*) T_{n-1} and T_n involved different stimuli but

TABLE 1

MEAN RTS (IN MSEC.) AND PERCENTAGE OF ERRORS (IN PARENTHESES) FOR THE THREE TYPES OF TRIAL PAIRS, SEPARATELY FOR EACH INTERTRIAL INTERVAL (ITI)

Trial pairs	ITI (in msec.)			
	100	500	1,000	Average
Type of transition				
I (Identical)	459 (1.0)	397 (3.0)	388 (1.7)	416 (1.9)
E (Equivalent)	527 (7.1)	448 (7.7)	437 (8.7)	472 (7.8)
D (Different)	547 (4.5)	453 (6.2)	438 (6.1)	479 (5.6)
Type of repetition effect				
E − I (stimulus)	68	51	49	56
D − E (response)	20	5	1	7

---

[3] Umiltá et al. (1972) have also discussed two hypotheses about the repetition effect and it is worth mentioning the relation between their hypotheses and ours. Umiltá et al. term their explanations the "memory and expectancy hypotheses." Their memory hypothesis corresponds to our trace decay hypothesis, except that Umiltá et al. do not specify that the memory trace of interest be a sensory or visual trace. Their expectancy hypothesis corresponds roughly to our memory search hypothesis, but Umiltá et al. consider only adjacent repetition effects in relation to the model.

[4] The reviewer of this paper has cleverly pointed out that our assignment of digits to responses resulted in a potential confounding similar to that present in the Bertelson (1965) study (see introduction). By assigning the digits 1, 2 to one response and 3, 4 to the other response, we made it possible for *S* to determine his response on the basis of whether the stimulus was low (1, 2) or high (3, 4). Had *S* used this strategy, there should have been no stimulus repetition effects in the present study; the fact that there were such effects indicates that *S* did not code the stimulus as low or high.

TABLE 2

Mean Reaction Times (in msec.), Percentage of Error (% E), and Number of Observations (N) for the Nine Types of Trial Triads, Separately for Each Intertrial Interval (ITI)

Type of trial triad[a]	Sample triad[b]	ITI (in msec.)											
		100			500			1,000			Average		
		RT	% E	N	RT	% E	N	RT	% E	N	RT	% E	N
II	111	440	0	404	369	1.1	352	363	.5	398	391	.5	1,154
EI	211	457	0	384	395	2.3	343	385	.5	376	413	.9	1,103
DI	311	471	2.1	755	412	4.2	727	407	2.9	759	430	3.1	2,241
IE	121	484	2.3	386	417	1.9	359	419	3.8	325	442	2.6	1,070
EE	221	540	9.5	373	449	12.3	335	444	11.8	351	479	11.2	1,059
DE	321	543	8.4	665	465	8.3	665	442	9.2	659	484	8.6	1,989
ID	131	537	4.4	625	450	4.4	634	433	5.6	628	473	4.8	1,887
ED	231	561	4.8	628	456	4.2	661	443	3.5	643	486	4.2	1,932
DD	331	540	4.4	1,454	452	7.8	1,402	438	7.5	1,409	477	6.6	4,265

[a] First letter gives transition between T_{n-2} and T_n; second letter gives transition between T_{n-1} and T_n; I = identical, E = equivalent, and D = different.

[b] In the sample triads, the test stimulus on T_n is always the digit 1, and the digits 1 and 2 are associated with the same response.

the same response (an equivalent [E] transition), or (c) T_{n-1} and T_n involved different stimuli and different responses (a different [D] transition). While the RT for each type of transition generally decreased as ITI increased (see Table 1), our main interest is in repetition effects. An estimate of the stimulus repetition effect is given by the E − I difference, presented in the fourth row of Table 1, and an estimate of the response repetition effect is given by the D − E difference, presented in the fifth row of Table 1. These repetition effects are not broken down by level of practice; although the stimulus repetition effect diminished with practice and the response repetition effect increased with practice (just as Rabbitt, 1968, found), neither of these practice effects reached significance. The results presented in Table 1 are directly comparable to those of Eichelman (1970). Like Eichelman, we found a sizable stimulus repetition effect ($p < .001$) which decreased as ITI lengthened ($p < .025$). Unlike Eichelman, we found a small response repetition effect ($p < .05$) which decreased as ITI lengthened ($p < .05$). But note that, at each ITI the faster RT at the E than D transition is accompanied by a higher error rate at the E than D transition; thus, the response repetition effect may be partly due to a speed–accuracy trade-off. So with regard to stimulus repetitions, our results are consistent with the Eichelman–Posner trace decay hypothesis, but a subsequent analysis will reveal the limitation of the present treatment of the data.

Before we move on to this more extensive analysis, there is an unexpected finding in the present results that deserves mention. If there was no stimulus repetition, latencies to the digits 2 and 3 were 32 msec. slower than latencies to the digits 1 and 4 ($p < .001$). This digit effect was greatly reduced, however, with stimulus repetitions. When a stimulus repetition occurred at the 100-msec. ITI, the digit effect disappeared; while when stimulus repetitions occurred at the 500- and 1,000-msec. ITIs, the digit effects were only 16 and 18 msec., respectively (though these effects are still significant [$p < .025$] in both cases). Moreover, error rates revealed the same types of digit effect. There are three possible explanations for this effect. First, the digits 1 and 4 may be more discriminable than the digits 2 and 3. Second, RTs to 1 and 4 may be faster because these digits are the "extremes" of the stimulus set. Finally, RTs to 2 and 3 may be slower because these digits lie at the boundary between responses. If either of the last two explanations is valid, then the topology of the

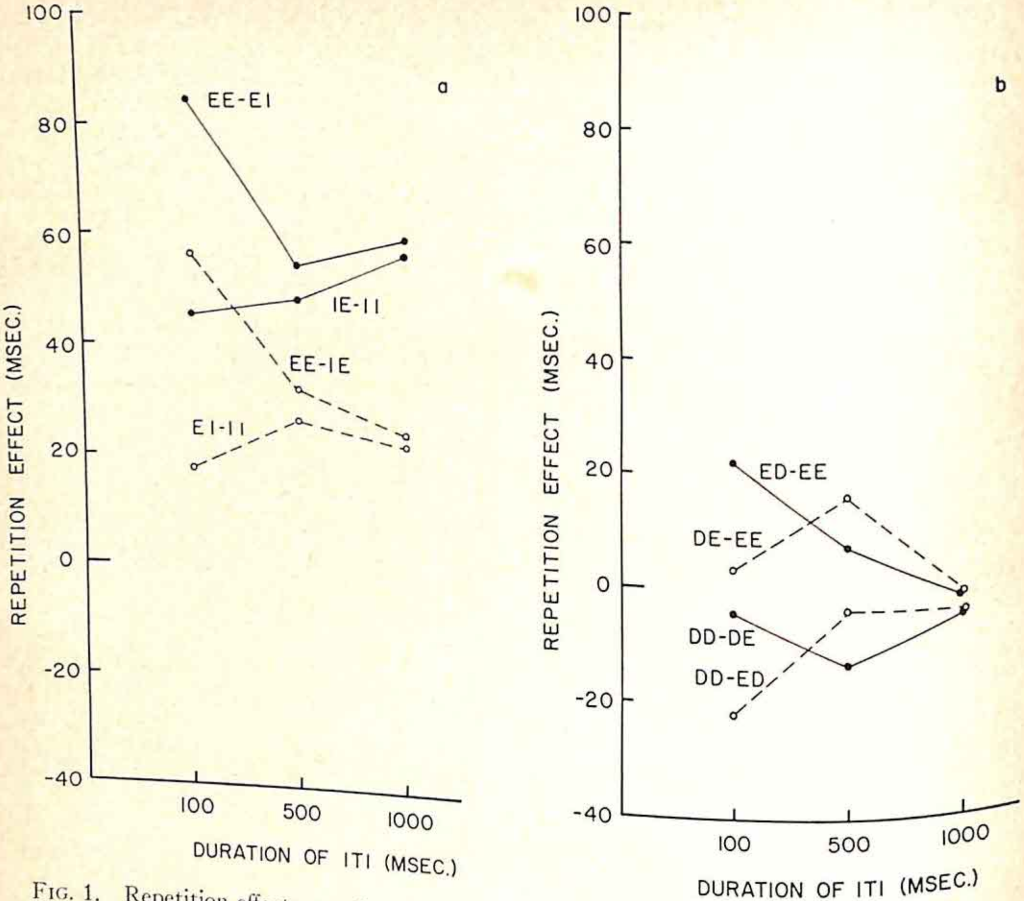


FIG. 1. Repetition effects as a function of the duration of the ITI. (Panel a contains the data for the two components of the adjacent stimulus repetition effect [EE − EI and IE − II] and the two components of the remote stimulus repetition effect [EE − IE and EI − II]; Panel b contains the data for the two components of the adjacent response repetition effect [ED − EE and DD − DE] and the two components of the remote response repetition effect [DE − EE and DD − ED].)

stimulus set is an important determinant of retrieval from STM, along with recency.

An analysis to include both adjacent and remote repetitions. The unit in this analysis is a triad of successive trials, T_{n-2}, T_{n-1}, and T_n, and the focus is on the effects of each of the preceding trials on the test trial (i.e., the effects of T_{n-2} on T_n and of T_{n-1} on T_n). Extending the symbolism previously used, each triad may be designated by a pair of letters (e.g., II), where the first letter specifies the remote transition between T_{n-2} and T_n, and the second letter specifies the adjacent transition between T_{n-1} and T_n. Since either a remote or an adjacent transition may be an I, E, or D type, there are now nine possible triads of interest. The nine rows of Table 2 present the data (number of observations, mean RT, and percentage of error) for each triad, separately for each ITI. There are two sets of four triads that are uniquely suited for evaluating stimulus and response repetition effects, since within one set, response effects are constant and stimulus effects vary, while the converse obtains in the other set. The first set consists of the II, EI, IE, and EE triads. Each of these triads contains both a remote and an adjacent response repetition, but within the set there is a factorial variation of remote stimulus repetitions (present or absent) with adjacent stimulus repetitions (present or absent). The set that will be used to analyze response effects contains the EE, DE, ED, and DD triads. None of these triads contains either a remote or an adjacent stimulus repetition, but within the set there is a factorial variation of remote and adjacent response repetitions.

Consider first stimulus repetitions. Two triads of interest contain an adjacent stimulus repetition, II and EI, and the other two triads in this set, IE and EE, differ from the above only in that they have an equivalent rather than an identical transition between T_{n-1} and T_n. Consequently there are two estimates or components of the adjacent stimulus repetition effect: IE − II and EE − EI. These components are presented in the solid lines of Figure 1a as a function of ITI. The two components manifest a substantial repetition effect ($p < .001$, in both cases), but the EE − EI component declines with ITI ($p < .025$), while the IE − EE component clearly does not (the slight increase over ITI for the latter component is insignificant). A similar pattern of results obtains when remote stimulus repetitions are considered. Two triads contain a remote repetition, II and IE, and the other two, EI and EE, differ only in that they have an equivalent transition between T_{n-2} and T_n. The two components of the remote repetition effect are, therefore, EI − II and EE − IE, and they are given by the dashed lines in Figure 1a. Again, the two components show a repetition effect ($p < .001$, in both cases), but one component (EE − IE) declines with ITI ($p < .005$) while the other (EI − II) does not.

Thus, for both remote and adjacent stimulus repetitions, the component of the repetition effect including an II triad (which is the only triad that contains two repetitions) does not decline with ITI, but the other components do. The results indicate that the previous Bertelson-type analysis was misleadingly simple in showing an adjacent stimulus repetition effect that declined with ITI. The previous analysis considered only the weighted average of what we are now calling components of the adjacent repetition effect, and it is now clear that this average does not characterize all components. Also, the Bertelson-type analysis included in its average a component which has not been considered here—DE − DI—since only one response repetition occurred in this component. It is worth mentioning, however, that the DE − DI component revealed a significant repetition effect ($p < .001$) which diminished with ITI ($p < .001$). This is not surprising as this component included triads manifesting at most one stimulus repetition, and the data in Figure 1a suggest that components fail to decline with ITI when they include triads manifesting two repetitions.

The same logic that was used with stimulus repetitions may be applied to response repetitions, this time using the triads EE, DE, ED, and DD. The two components of the adjacent repetition effect, ED − EE and DD − DE, are presented in the solid lines of Figure 1b, while the two components of the remote repetition effect, DE − EE and DD − ED, are given by the dashed lines of Figure 1b. There is neither a remote nor an adjacent response repetition effect present. These results conflict with that obtained from the earlier Bertelson-type analysis, namely, a significant adjacent response repetition effect that declined with ITI. This conflict can be resolved by recalling that the earlier analysis considered the weighted average of the two components in Figure 1b and the component ID − IE. The latter component was not considered in the present analysis since, unlike the other two components of the adjacent response repetition effect, the ID − IE component includes triads containing a stimulus repetition. This component, however, produced a significant adjacent response repetition effect ($p < .001$) that diminished with ITI ($p < .001$), and this was the sole cause of the results obtained with the Bertelson-type analysis. The fact that an adjacent response repetition effect occurred only when there was a stimulus repetition present suggests an interaction between stimulus and response repetition effects. Such an interaction in turn implies a problem with our assumptions that: (*a*) The stages underlying

STM retrieval can be characterized as stimulus or response stages and (*b*) stimulus and response repetitions affect only stimulus and response stages, respectively (D. Massaro, personal communication, June 1972). This problem will be treated in the Discussion section.

In any event, response repetition effects, as measured by the present means, are relatively trivial when compared to stimulus repetition effects. This conclusion is consistent with that emerging from other studies of Bertelson's (1965) repetition effect (e.g., M. C. Smith, 1968; see introduction). It is also consistent with most of the best executed studies on the effects of probability or expectancy on STM retrieval which consistently show that stimulus probability or expectancy affects retrieval time more than response probability or expectancy (Biederman & Zachary, 1970; Hinrichs & Krainz, 1970).

DISCUSSION

Trace decay vs. memory search. The major findings of our extended analysis of repetition effects are: (*a*) Both adjacent and remote stimulus repetition effects occur and behave similarly with respect to ITI variations, (*b*) some components of stimulus repetition effects decline with ITI while others do not, and (*c*) compared to stimulus repetition effects, response repetition effects are minor. Our major concern is to use these findings, and others, to distinguish between the trace decay and memory search hypotheses.

Consider first the trace decay hypothesis in relation to the present data. The only strong evidence for a trace decay hypothesis of the Eichelman–Posner variety is the decline of some components of the adjacent repetition effect with ITI. But our results show this same ITI finding for components of the remote repetition effect, a repetition effect which is unlikely to occur if trace decay is operative. Since the ITI finding is as likely to be manifested with a repetition effect not predicted by trace decay as it is with a repetition effect that is predicted by this theory, it seems that the ITI finding can no longer be taken as evidence for this hypothesis. Thus, there is little to recommend the trace decay explanation.

The memory search hypothesis, however, does appear to be quite viable when judged against the current data. This hypothesis predicts both remote and adjacent stimulus repetition effects, with the former being of lesser magnitude than the latter. All of these predictions were confirmed. Further, this hypothesis can also accommodate a decrease in stimulus repetition effects with ITI if it is assumed that task-irrelevant stimuli are increasingly likely to enter STM as the ITI lengthens. To explain why components of the repetition effect involving triads with two repetitions (II triads) do not decline with ITI, it must be further assumed that a twice repeated stimulus prevents *S* from attending to task-irrelevant stimuli. But this last assumption is totally post hoc, and the fact that different components of the repetition effect are differentially affected by ITI remains a problem for any theory of repetition effects.

One potential problem with the above version of the memory search hypothesis is the assumption that repetition effects decline with ITI because task-irrelevant stimuli are more likely to enter STM. If this were indeed the case, then the representations of all the stimulus alternatives in STM should be lowered in the search order as ITI lengthens, and one might have expected overall RT to increase with ITI. In fact, overall RT decreased with ITI (see Tables 1 and 2). This increase may turn out to be independent of variations in search order, but for now, it seems prudent to consider two other versions of the memory search hypothesis which differ from the above in how they explain decreases in repetition effects with lengthening ITI.

The first alternative version to be considered emphasizes rehearsal. Suppose that immediately after the presentation of an item, the search order of the memorial alternatives is probabilistically determined by the recency of these alternatives, but as the time until the presentation of the next item (i.e., the ITI) increases there is an increasing chance that *S* will rehearse one or more of the memorial alternatives and their response assignments. If a rehearsal functions like a presentation, then the rehearsed alternative will, with some probability, move up in the search order. This offers a means by which remote and adjacent stimulus repetition effects might decline with ITI. Specifically, longer ITIs leave more time for rehearsal, which results in rehearsal becoming a determinant of search order. If the rehearsed alternative differs from the most recently presented ones, then the net effect will be to reduce the influence of recency on search order and hence reduce repetition effects. To explain why components involving triads with two repetitions do not decline with ITI, an ad hoc assumption is again required, i.e., it must be assumed that when a stimulus has just occurred twice it is extremely likely to be rehearsed.

Consider now the second way in which a memory search hypothesis can be modified so as to be compatible with repetition effects that decrease over ITI. It is possible that as the ITI lengthens there is a change from a self-terminating memory search to a more exhaustive one, i.e., search continues after the target alternative has been located. Sternberg (1969) has consistently found some evidence for an exhaustive search of STM in his paradigm (which differs somewhat from the present one), and Sternberg has consistently used ITIs and memory set – test stimulus intervals that are longer than the longest ITI in the present study (1,000 msec.). In fact, when the memory set – test stimulus interval in the Sternberg task is shortened to 800 msec., a recency effect occurs (Clifton & Birenbaum, 1970). A transition from a self-terminating to a more exhaustive search would also explain why overall RT decreases as ITI lengthens, as Sternberg (1969) has argued that exhaustiveness leads to faster searching than self-termination and Clifton and Birenbaum (1970) have provided some support for this contention. However, there are difficulties with this self-termination vs. exhaustiveness explanation of the ITI effect. First, whether memory search in the Sternberg paradigm is really exhaustive is itself a matter of current debate (Theios et al., 1973). Second, several studies (e.g., Schvaneveldt & Chase, 1969) have found substantial repetition effects at ITIs which are even longer than the intervals used by Sternberg (1969), and such intervals should presumably lead to relatively exhaustive searching and negligible repetition effects.

Stimulus vs. response stages and repetition effects. We noted earlier (see Results section) a problem with our assumptions that STM retrieval stages can be characterized as stimulus or response stages and that stimulus and response repetitions affect these stages selectively. The problem was that there was evidence for an adjacent response repetition effect only when a stimulus repetition also occurred. This unexpected result suggests that our assumptions may be in need of revision, and the following offers a plausible reinterpretation.

While we have assumed (see introduction) that our task requires four distinct processing stages—(*a*) stimulus encoding, (*b*) stimulus comparison, (*c*) response selection, and (*d*) response execution—it is quite possible that the second and third stages were combined into a single stage in the present experiment. That is, what *S* may have stored in STM is a representation of each stimulus *marked* with an indication of its appropriate response. If this were the case, then the second or comparison stage would involve a search of stimulus representations *and* their response markings, and an output of this stage would be an indication of the selected response. So, stimulus comparison and response selection would have been collapsed into a single stage. Now, it is no longer puzzling that stimulus and response repetition effects might interact, as they could both effect the combined stage.

Given this three-stage viewpoint, what is the status of our findings on stimulus and response repetitions, and the interpretation of our results, in terms of memory search? One can still make a case that the response-execution stage is the only stage affected by response repetitions. Consequently, the minor response effects obtained imply that this stage makes little contribution to Bertelson's (1961) repetition effect. The fact that stimulus repetition effects were far greater than their response counterparts may be interpreted as showing that the present stimulus factors were the more important ones in determining the duration of the combined stage. And the memory search hypothesis still stands; all that has changed is what gets searched—stimulus representations *plus* response markings, rather than only stimulus representations.

One final point about a sequential stage analysis deserves mention. This article, like several previous ones, has assumed that the locus of the stimulus repetition effect is at the comparison (or combined) stage of STM retrieval, rather than the prior encoding stage. But this assumption has yet to receive any direct experimental support. Moreover, it is well known that stimulus repetitions can facilitate encoding when tachistoscopic presentations are employed (Doherty & Keeley, 1972). Clearly, further attempts at a theoretical explanation of sequential RT effects must consider this encoding vs. comparison issue.

REFERENCES

Bertelson, P. Sequential redundancy and speed in a serial two-choice responding task. *Quarterly Journal of Experimental Psychology*, 1961, **13**, 90–102.

BERTELSON, P. Serial choice reaction-time as a function of response versus signal- and response-repetition. *Nature*, 1965, **206**, 217–218.

BIEDERMAN, I., & ZACHARY, R. A. Stimulus versus response probability effects in choice reaction time. *Perception & Psychophysics*, 1970, **7**, 189–192.

CLIFTON, C., JR., & BIRENBAUM, S. Effects of serial position and delay of probe in a memory scan task. *Journal of Experimental Psychology*, 1970, **86**, 69–76.

DOHERTY, M. E., & KEELEY, S. M. On the identification of repeatedly presented, brief visual stimuli. *Psychological Bulletin*, 1972, **78**, 142–154.

EICHELMAN, W. H. Stimulus and response repetition effects for naming letters at two response-stimulus intervals. *Perception & Psychophysics*, 1970, **7**, 94–96.

HINRICHS, J. V., & KRAINZ, P. L. Expectancy in choice reaction time: Anticipation of stimulus or response. *Journal of Experimental Psychology*, 1970, **85**, 330–334.

KORNBLUM, S. Sequential determinants of information processing in serial and discrete choice reaction time. *Psychological Review*, 1969, **76**, 113–131.

POSNER, M. I., BOIES, S. J., EICHELMAN, W. H., & TAYLOR, R. L. Retention of visual and name codes of single letters. *Journal of Experimental Psychology Monograph*, 1969, **79** (1, Pt. 2).

RABBITT, P. M. A. Repetition effects and signal classification strategies in serial choice-response tasks. *Quarterly Journal of Experimental Psychology*, 1968, **20**, 232–240.

SCHVANEVELDT, R. W., & CHASE, W. G. Sequential effects in choice reaction time. *Journal of Experimental Psychology*, 1969, **80**, 1–8.

SMITH, E. E. Choice reaction time: An analysis of the major theoretical positions. *Psychological Bulletin*, 1968, **69**, 77–110.

SMITH, M. C. The repetition effect and short-term memory. *Journal of Experimental Psychology*, 1968, **77**, 435–439.

STERNBERG, S. Memory scanning: Mental processes revealed by reaction-time experiments. *American Scientist*, 1969, **57**, 421–457.

THEIOS, J., SMITH, P. G., HAVILAND, S. E., TRAUPMANN, J., & MOY, M. C. Memory scanning as a serial self-terminating process. *Journal of Experimental Psychology*, 1973, **97**, 323–336.

UMILTÁ, C., SNYDER, C., & SNYDER, M. Repetition effect as a function of event uncertainty, response-stimulus interval, and rank order of the event. *Journal of Experimental Psychology*, 1972, **93**, 320–326.

(Received September 2, 1972)

Journal of Experimental Psychology
1973, Vol. 98, No. 2, 423–425

SHORT REPORTS

RELEASE FROM PROACTIVE INTERFERENCE IN THE RECALL OF SENTENCES[1]

RICHARD E. SCHUBERTH
Connecticut College

BARRY L. LIVELY
Richard H. Hutchings Psychiatric Center

DONALD B. REUTENER[2]
Smith College

Two short-term memory experiments were performed using the Peterson and Peterson distractor technique. The materials to be recalled were sentences. In Experiment I, after nine trials, shifts were made in either or both the nature of the sentence frame (declarative or interrogative) and the semantic categories of critical words (subject, direct object, and locative phrase). A control group in which no shifts were made was also included. In Experiment II, only characteristics of the sentence frame were changed. It was found that retention of critical words declined over preshift trials in both experiments and that only shifts in semantic categories of words produced an improvement in performance.

Wickens (1970) treats release from proactive interference (PI) in the Peterson and Peterson (1959) short-term memory (STM) paradigm as an empirical means for determining the way in which items are encoded in memory. Loess (1967) observed a reduction in recall of word triads drawn from the same semantic category (birds, professions, etc.) as a function of trials with improvement subsequent upon a change in category. The improvement in performance (release from PI) suggests that the encoding of the words presented reflected the difference in semantic category.

Experiments have also been reported in which syntactic characteristics were involved in the shift. Wickens, Clark, Hill, and Wittlinger (1968) presented word triads drawn from a set of verbs or adjectives for four trials followed by a shift to the other class. No release from PI was observed. It would be premature, on the basis of these results, to argue that syntactic category is not represented in the encoded form of the to-be-remembered material; it simply may be that syntactic coding is not effective outside the context of sentences. Heisey and Duncan (1971) had *S*s recall nonsense-word triads or sentences where nonsense words served as subject, verb, and object. The sentences in the first three trials were in the active, passive, or interrogative voice; on Trial 4, all *S*s were given nonsense sentences in the active voice. The shift from nonsense-word triads produced a large release from PI, while the shifts from passive and interrogative produced only a moderate release.

Two experiments are reported here in which semantic and syntactic characteristics of real sentences were manipulated. In both experiments a basic sentence frame of the form "____ did buy a ____ in ____" was used. In Experiment I, a semantic shift was achieved by changing the semantic categories of subject, direct object, and locative phrase. A syntactic shift involved a change in sentence form (declarative vs. interrogative). This particular syntactic shift was selected because it necessitated only a minimum change in word order, a variable which could conceivably produce a release from PI. In Experiment II, only characteristics of the sentence frame were shifted. The meaning was varied by shifting between affirmative and negative sentences, while a syntactic shift involved changing the position of the locative in the sentence. In both experiments, different groups were exposed to shifts in semantic, syntactic, or semantic and syntactic characteristics. Control groups were run in which no shift was employed.

Method. The *S*s in Experiment I were 61 male and 3 female students solicited from the population at Bowdoin College. The *S*s in Experiment II were 31 male and 1 female students fulfilling an introductory psychology course requirement at Bowdoin College. In both experiments, *S*s were assigned randomly to a condition upon their arrival at the laboratory.

Materials in both experiments were presented by a Kodak Carousel projector governed by a Gerbrands tape timer. Each *S* was given 12 trials. On each trial, *S* was shown a sentence (to be read aloud) for 2 sec. followed by a 6 × 8 matrix of random digits, the digits of each row to be added consecutively and aloud for 16 sec. Presentation of a blank slide signaled the 10-sec. recall interval. An asterisk presented for 2 sec. indicated the beginning of the next trial. A practice trial was given where the letters ABCD were used as the to-be-remembered item.

The materials used in Experiment I were parallel sets of active and interrogative sentences which were constructed by inserting one of two sets of

---

[1] This research was supported in part by National Science Foundation Institutional Grant GU-3444 to Smith College.

[2] Requests for reprints should be sent to Donald B. Reutener, Department of Psychology, Smith College, Northampton, Massachusetts 01060.

critical words into the frames: (*a*) "____ did buy a ____ in ____." and (*b*) "Did ____ buy a ____ in ____." One set of critical words was selected from the 25 most frequent responses (Battig & Montague, 1969) to Word Categories 47 (a male's first name), 17 (a weapon), and 54 (a city); the second set was selected from Categories 46 (a girl's first name), 19 (a type of human dwelling), and 21 (a country). Twelve sentences, each constituting a trial, were constructed to produce four sets of sentences. The 12 trials were treated as four blocks of three trials. In the first three blocks, experimental *S*s were exposed to sentences in which the semantic categories of the subject, direct object, and locative phrase remained the same and the syntactic structure was constant. The shift conditions were achieved in the fourth block by changing either: (*a*) the semantic categories of the three critical words, (*b*) the structure of the sentence frame, or (*c*) both categories and structure. Control *S*s were exposed throughout all four trial blocks to sentences corresponding in both structure and categories to those received by all experimental *S*s on the fourth trial block. Sentences were rotated across trial blocks in each group so that each sentence appeared in each of three positions an equal number of times. Trial blocks were also rotated so that each trial block appeared as either Block 1, 2, 3, or 4 an equal number of times.

In Experiment II, parallel sets of sentences of the form "____ did buy a ____ in ____."; "In ____, ____ did buy a ____."; "____ can't buy a ____ in ____."; and "In ____, ____ can't buy a ____." were constructed in which the subject, direct object, and locative phrase were the same as those used in Experiment I from Battig and Montague (1969) Categories 46, 19, and 21.

A shift in the position of the locative phrase was labeled a *syntactic word order* shift, while the affirmative–negative change was called a *semantic frame* shift. The *S*s were treated in the same manner as in Experiment I with different groups of *S*s given shifts appropriate to the requirements of the design: syntactic word order, semantic frame, semantic frame–syntactic word order, and control.

Results. The dependent variable in both experiments was the number of critical items recalled in their correct locations within each trial block. Thus, the *S*'s score could range from zero to nine for each trial block.

TABLE 1

PROPORTION OF CORRECT RESPONSES AS A FUNCTION OF NATURE OF SHIFT AND TRIAL BLOCK, EXPERIMENT I

Nature of shift	Trial block			
	1	2	3	4
Control	.83	.58	.54	.45
Semantic	.85	.66	.61	.80
Syntactic	.88	.62	.54	.43
Semantic–Syntactic	.85	.61	.55	.71

TABLE 2

PROPORTION OF CORRECT RESPONSES AS A FUNCTION OF NATURE OF SHIFT AND TRIAL BLOCK, EXPERIMENT II

Nature of shift	Trial block			
	1	2	3	4
Control	.86	.71	.65	.68
Semantic frame	.88	.71	.68	.64
Syntactic word order	.83	.68	.65	.60
Semantic frame–Syntactic word order	.71	.57	.42	.44

Visual inspection of Table 1 suggests that control and experimental groups in Experiment I showed equal decrements in performance over the first three trial blocks, and that an increment in performance was obtained on the fourth trial block only for the semantic and semantic–syntactic shift groups.

The results of an analysis of variance showed that performance over the first three trial blocks did not differ among groups, $F\ (3,\ 48) < 1$, indicating comparability of groups prior to the trial block in which the shift was made. Performance declined significantly over the first three trial blocks, $F\ (2,\ 96) = 46.85$, $p < .01$, a result predicted on the basis of a PI buildup.

An analysis of variance for performance on the third vs. the fourth trial block for all groups revealed a significant Blocks × Groups interaction, $F\ (3,\ 48) = 6.12$, $p < .01$. Separate analyses of variance for the individual groups showed that the change in performance from the third to the fourth trial block was not significant for either the control group, $F\ (1,\ 12) = 1.49$, $p > .05$, or the syntactic shift group, $F\ (1,\ 12) = 3.88$, $p > .05$, but was significant for both the semantic shift group, $F\ (1,\ 12) = 9.56$, $p < .01$, and the semantic–syntactic shift group, $F\ (1,\ 12) = 5.86$, $p < .05$.

The results of Experiment II are displayed in Table 2, where equal decrements in performance over the first three trial blocks are seen for control and experimental groups. A slight increment in performance was obtained on the fourth trial block for the control and semantic frame–syntactic word order shift groups, but not for the semantic frame or syntactic word order shift groups.

Overall performance for the groups was found to be significantly different, $F\ (3,\ 24) = 3.12$, $p < .05$, due mainly to the depressed performance of the semantic frame–syntactic word order shift group. This difference in performance is seen to persist over trial blocks; since the groups were exposed to the same materials and were treated in the same way prior to the shift trial, this must represent a subject-selection factor. Performance for all groups over the four blocks declined significantly, $F\ (3,\ 72) = 9.31$, $p < .01$, a result predicted on the basis of PI buildup. The Groups × Blocks interaction was not significant, $F\ (9,\ 72) < 1$. The data for all groups were compared on the third and fourth trial blocks to determine whether the small divergence in performance noted for the control and semantic frame–syntactic word order shift groups vs. the semantic frame and syntactic word order shift groups was significant. The Groups × Blocks interaction was not significant, $F(3, 24) < 1$.

Discussion. In the present work, as in the literature cited by Wickens (1970), PI was found to build up as a function of trials on similar material. A subsequent release from PI for a set of critical words drawn from different semantic categories was observed, while a release was not found when the order of the critical words changed or shifts occurred in syntactic or semantic relations elsewhere in the sentence.

The failure to show an effect of shift in syntax stands in apparent contrast to the results of Heisey and Duncan (1971). The studies were different in several ways however, including the type of materials used, number of preshift trials, and length of stimulus presentation. An attempt at resolving these differences is beyond the scope of this article. It should be noted, however, that Heisey and Duncan report only a moderate release when the shift is accomplished according to voice.

It may be that while grammatical distinctions are important for the generation and interpretation of utterances, the STM distractor technique is not sensitive to variations in syntax. This could result from the rote nature of the task. The *S* is not required to give any discriminative response to a sentence which may tap the meaning of the sentence. That is, *S* is not required to indicate agreement or disagreement with the utterance; he is only required to reproduce the sentence after some interval of activity. If *S* were required in some way to attend to the "meaning" of a sentence, changes in syntax might produce a release from PI. A second possibility is that the change in structure was not sufficient to produce a release from PI. A greater change in structure is almost necessarily accompanied by a drastic change in word order and the importation of new words into the sentence frame. Since either of these variables might produce a release from PI, future work employing greater structural changes will require a careful assessment of the effects of possible artifactual effects due to these variables.

REFERENCES

Battig, W. F., & Montague, W. E. Category norms for verbal items in 56 categories: A replication and extension of the Connecticut category norms. *Journal of Experimental Psychology Monograph*, 1969, **80** (3, Pt. 2).

Heisey, J. A., & Duncan, C. P. Syntactical encoding in short-term memory. *Journal of Verbal Learning and Verbal Behavior*, 1971, **10**, 95–100.

Loess, H. Short-term memory, word class, and sequence of items. *Journal of Experimental Psychology*, 1967, **74**, 556–561.

Peterson, L. R., & Peterson, M. J. Short-term retention of individual verbal items. *Journal of Experimental Psychology*, 1959, **54**, 157–173.

Wickens, D. D. Encoding categories of words: An empirical approach to meaning. *Psychological Review*, 1970, **77**, 1–15.

Wickens, D. D., Clark, S. E., Hill, F. A., & Wittlinger, R. P. Investigation of grammatical class as an encoding category in short-term memory. *Journal of Experimental Psychology*, 1968, **78**, 599–604.

(Received January 26, 1972)

Journal of Experimental Psychology
1973, Vol. 98, No. 2, 426–428

COREFERENCE AND SHORT-TERM MEMORY FOR DISCOURSE[1]

ROBERT J. JARVELLA[2]

Case Western Reserve University

Listeners' recall and recognition were obtained for either the final or next-to-final sentence heard in 100–200 word segments of recorded dialogue. In each task, response accuracy was significantly greater for certain phrases from test sentences when the following context repeated or reimplied the phrases than when it was nonredundant. The phrases were best remembered when repeated, but the remaining portions of the sentences were best remembered when the phrases were only reimplied. The content of successive sentences and clauses of heard speech appears to influence the forgetting of their verbatim form.

Short-term memory (STM) for connected speech has recently been studied by having *S*s try to recall some of the end of a narrative passage which they have just heard (Jarvella, 1970, 1971, 1972; Jarvella & Herman, 1972). Two principal findings from this research have been that (*a*) recall is typically quite accurate for only the final sentence of discourse heard, and (*b*) recall is usually perfectly verbatim for only the most recent clause from this final sentence. These results suggest that as a clause or sentence is succeeded by another such unit, its representation in STM is often partly erased. The main question raised by the present research concerns whether such apparent forgetting is partly a function of the *content* of the following sentence or clause heard.

When two phrases have the same content, they are usually said to be coreferential. The grammatical form of coreference is often abbreviated. Some familiar examples of this phenomenon include the replacement of a noun phrase by a pronoun or the deletion of a redundant verb phrase or complement. For example, we understand the sentence, "The boy read the book but the girl did not," to mean that the boy read the book but that the girl did not *read the book*. Since successive sentences and clauses may be partly coreferential or noncoreferential, grammatical structure and propositional content at these levels can be varied somewhat independently.

In the present study, coreference was varied between a previously heard sentence or clause and a following final one; one of the final two sentences was then cued for recall or recognition. It was found that a phrase from the previous sentence or clause could be quite accurately remembered if it was coreferential with (had the same content as) part of the following such grammatical unit. Full and grammatically abbreviated forms of coreference, however, yielded somewhat different results. While the test sentences' coreferential phrases were best remembered with full repetition, their remaining noncoreferential portions were remembered best when the coreference was only anaphoric. These new results are discussed in relation to a previously suggested theory of short-term storage of heard speech (Jarvella, 1971).

Method. Certain phrases embedded in discourse were followed by a largely coreferential or noncoreferential context. Three kinds of contexts were employed, one repeating the phrase fully (Repeated), a second deleting the phrase but still reimplying it (Implied), and a third including only generally new information (Novel). The phrase occurred as part of one of the final two sentences presented (previous or immediate); a short new sentence or a further clause of the same sentence made up the following context. After the context was presented, the full sentence containing the phrase was prompted for recall or recognition.

The following examples include 1 of 12 sets of test items which were developed for each sentence position tested. The phrases shown in italics were coreferential only in Implied and Repeated; these are later referred to as the sentences' critical phrases and distinguished from their remaining noncritical portions.

Previous sentence prompted: "I suppose *the reflected headlights probably blinded him*. Don't you think so?" (Implied); ". . . Don't you think the reflected headlights probably blinded him?" (Repeated); ". . . Weren't we both lucky?" (Novel).

Immediate sentence prompted: "I want to *hear this saddening account detailed*, and I will, too." (Implied); ". . ., and I will hear this saddening account detailed, too." (Repeated); ". . ., before something tragic happens." (Novel).

The previous sentences used were identical across context conditions, averaging about 11 words long; in Novel and Implied, the following context averaged about 4 words vs. 9–10 words in Repeated. Similarly, immediate sentences averaged about 14 words in length in Novel and Implied but 19–20 words in Repeated. Thus, Novel and Implied were always matched in length, and Repeated was longer

---

[1] This article is based in part on a paper presented at the meeting of the Acoustical Society of America, Denver, October 1971. This research was supported in part by Grant DAHC 15 to Rockefeller University from the Advanced Research Projects Agency, and Grant MH-22134 to Case Western Reserve University from the National Institute of Mental Health. The author wishes to thank George A. Miller and Sheldon Rosenberg for contributing helpful criticism.

[2] Requests for reprints should be sent to Robert J. Jarvella, Department of Psychology, Case Western Reserve University, Cleveland, Ohio 44106.

TABLE 1

Percentage of Correct Recall and Recognition

Unit	Previous sentence			Immediate sentence		
	Implied	Repeated	Novel	Implied	Repeated	Novel
Recall						
Sentence	42	51	26	36	20	11
Critical phrase	60	78	45	57	70	28
Noncritical portion	66	65	61	64	53	44
Recognition						
Same (nothing changed)	96	82	76	100	78	89
Different (critical phrase changed)						
Full	42	46	30	33	28	11
Full + Part	56	59	45	47	56	36
Different (noncritical portion changed)						
Full	35	27	26	45	29	27
Full + Part	53	43	43	68	45	45

than both by the number of words (5 or 6) in the repeated phrase.

The test items were placed at the end of 24 segments of dialogue adapted from radio plays. These segments ranged about 100–200 words in length. Three versions of the dialogues were constructed, each containing four instances of each sentence position (previous and immediate) by context condition (Implied, Repeated, and Novel) in a balanced order. Each version, together with instructions to *S*s, was tape recorded by a man and a woman speaking at about five syllables (three words) per second. Tone prompts were dubbed in just after each segment, at 200 Hz. for previous sentences and 600 Hz. for immediate ones.

The *S*s volunteering for the experiment were enrolled in psychology courses at Case Western Reserve University. They were tested in groups of 12 or 13 *S*s each. One recall and two recognition groups were randomly assigned to hear each version of the dialogues. Thus, nine different groups of *S*s were tested in all. All *S*s were instructed to attend to the conversations normally and not try to anticipate which sentences would be prompted. The prompting procedure (previous = low tone, immediate = high tone) was explained, demonstrated through examples, and displayed throughout the experiment on a blackboard. Each time a prompt sounded during the experiment, *E* stopped the tape recorder playing the dialogues until all *S*s had written their responses.

The recall *S*s' task was to write each prompted sentence down verbatim on a separate blank page in a test booklet. The recognition *S*s' task was to decide whether the prompted sentences were the same as or different from sentences printed on alternating pages of test booklets. On hearing a prompt, the recognition *S*s turned to the page with the printed sentence and gave three responses: (*a*) a same–different judgment on the prompted and printed sentences, (*b*) an identification of any change recognized in the printed version, and (*c*) a confidence rating (very confident, confident, not confident) on the same–different judgment. Of the 24 sets of recognition items, 4 were printed unchanged and the remaining 20 were printed with a slight change made in either the test sentence's critical or noncritical portion. These changes repeated the most frequent error in each sentence portion made by the recall *S*s. Usually this involved substitution, deletion, or addition of one word.

Results. In general, the data obtained for recall and recognition of the prompted sentences were quite comparable; they are presented together in Table 1. For recall, the percentages given in Table 1 are for sentences and parts of sentences which were written down verbatim, or exactly as presented. For recognition, the percentages given reflect correct same and different judgments. Of the two recognition measures shown for changed sentences, the full measure includes only responses where *S* marked different and also identified the change made; the full + part measure further includes different judgments where *S* did not fully identify the change. Response confidence was found to be closely related to recognition accuracy; correct responses were more confident than errors, and fully correct different judgments were more confident than partly correct ones ($p < .01$, for each by Wilcoxon matched-pairs signed-ranks tests by items).

The results for previous and immediate sentences are presented below separately. It will be seen that the differences obtained among the context conditions (Implied, Repeated, and Novel) were quite similar for each sentence position. Wilcoxon tests for matched pairs of items were used to obtain the significance levels reported.

For previous sentences, significantly poorer recall and recognition were generally obtained when the test sentence and following context were not coreferential. In Novel, fewer critical phrases and full sentences were recalled and fewer changes in critical phrases recognized than in either Repeated or Implied ($p < .01$, in all cases). Second, more full critical phrases from previous sentences were recalled with full than implied repetition ($p < .025$), and slightly more changes in these phrases were recognized. And third, more changes in the noncritical portions from previous sentences were recognized in Implied than in either Repeated or Novel ($p < .05$, for each), and slightly more of these portions were recalled.

For immediate sentences, again the poorest performance on most measures was obtained with the noncoreferential context. In Novel, fewer immediate sentence critical phrases were recalled, and fewer changes in these phrases were recognized than in either Repeated or Implied ($p < .025$, in all cases). Second, again relatively more critical phrases were recalled and recognized as being changed with full then implied repetition ($p < .05$, for recall and full + part recognition measure). Third, more full sentences and their noncritical portions were recalled and more changes in these portions were recognized in Implied than in either Repeated or Novel ($p < .05$, in all cases).

Finally, the difference in correct same judgments favoring Implied over Repeated and Novel was also significant over the four sets of items left unchanged in the recognition task ($p < .05$, by Friedman one-way analysis of variance by items).

Discussion. Previous studies of STM for discourse have found that listeners' recall for the most recent speech heard tends to be modulated by its grammatical structure: Usually only the final clause and its larger sentence are remembered very accurately. This finding suggests that processing of the discourse is guided by a structural analysis which results in temporary storage of one sentence and one clause at a time (Jarvella, 1971). The present study was conducted largely to determine whether the forgetting of these sentence and clause units is partly a function of successive units having distinct structural meanings. Two results bearing on this question will be considered.

First, it was found that certain phrases from a previously heard constituent were remembered better when they were reimplied in the following speech context than when this speech ended nonredundantly. The relatively high level of accuracy obtained with the abbreviated form of coreference suggests that in some cases sentence and clause units are not erased from STM as further speech is heard. It seems plausible, moreover, that accurate memory for a previously heard constituent may aid the listener to fill in information which is only implied in the following speech. In this regard, Bever (1970) has proposed a number of processing strategies which a listener might follow in recovering the structural meaning of a sentence from its surface form. On the other hand, whenever successive sentences or clauses have more fully distinct informational content, they may be processed more on a unit-by-unit basis. The relatively large degree of forgetting which was obtained with the Novel context appears to support this hypothesis.

Second, somewhat different patterns of results were obtained using fully redundant and linguistically shortened forms of coreference. In general, overt repetition led to better recall and recognition of the repeated material while the abbreviated form led to more uniformly high accuracy throughout the sentence being tested. These findings may further support the listening processes suggested above by the remembering of surface grammatical structure. It would seem just as feasible for the processing of discourse to proceed sentence by sentence and clause by clause if the speech were fully redundant (coreference was always explicit) as if it were fully nonredundant: In either case, the lexical information needed to relate a constituent to its structural meaning would fall within the constituent proper. If, in fact, fully redundant speech is processed in this way, then memory for the repeated phrases in the present study may be attributed mainly to their dual occurrence rather than to retention of the previous sentence or clause containing them. The reciprocal advantage obtained with the abbreviated form of coreference suggests this was the case, although this effect may have been due to shorter stimulus items in the Implied than the Repeated condition. A similar result favoring Implied, however, was also obtained with respect to the Novel context with which it was matched in length. Perhaps a more general explanation for these findings is that the processing of a fully explicit clause or sentence tends most to erase the previous unit from STM.

REFERENCES

Bever, T. G. The cognitive basis for linguistic structures. In J. R. Hayes (Ed.), *Cognition and the development of language.* New York: Wiley, 1970.

Jarvella, R. J. Effects of syntax on running memory span for connected discourse. *Psychonomic Science*, 1970, **19**, 235–236.

Jarvella, R. J. Syntactic processing of connected speech. *Journal of Verbal Learning and Verbal Behavior*, 1971, **10**, 409–416.

Jarvella, R. J. Speech processing memory. Paper presented at the International Conference on Speech Communication and Processing, Boston, April 1972.

Jarvella, R. J., & Herman, S. J. Clause structure of sentences and speech processing. *Perception & Psychophysics*, 1972, **11**, 381–384.

(Received May 20, 1972)

Journal of Experimental Psychology
1973, Vol. 98, No. 2, 429–431

CUED AND UNCUED FREE RECALL OF UNRELATED WORDS FOLLOWING INTERPOLATED LEARNING[1]

DAVID R. BASDEN[2]

California State University at Fresno

The *S*s were given free-recall training to criterion with a list of 30 unrelated words, original learning (OL), were given seven trials with a second list, and were then tested for OL retention. Half of the *S*s were cued with list members taken from alternate positions of the final recall of OL, half were not cued. Cued *S*s retained significantly more critical items than did control *S*s. Although similar results have been offered in support of the trace-dependent storage view, it was argued that such studies do not provide a differential test of independent vs. dependent storage.

In two recent papers (Slamecka, 1968, 1969), it has been argued that in free-recall learning, memory traces for the words of the list are stored independently. Organization appearing in recall order is said to result from the use of an organized retrieval plan. In addition, the relative efficiency of such organized retrieval plans is assumed to produce the heightened recall usually associated with increases in recall organization. The alternative position is that organization takes place at the time of storage; i.e., that traces of list members are stored in some dependent fashion. According to this view, recall of one member of a set of relationally stored items (S unit) is assumed to cue recall of other items from the set. Thus, this trace-dependent storage hypothesis attributes both organization in recall order and concomitant increases in recall to organization in storage.

Slamecka (1968, 1969) has argued that if, as indicated by the trace-dependent storage hypothesis, recall of one member of an S unit cues recall of the remaining members of that unit, then providing a member of that unit at the time of recall should have a similar effect. In a series of six experiments, Slamecka (1968) compared recall when some of the list members were provided as cues with recall when no such cues were provided. In none of these was cued recall found to be superior to uncued recall. In a subsequent series of experiments, Slamecka (1969) presented lists of unrelated words in constant serial order, cuing half of the *S*s with words taken from odd-numbered serial positions in the presentation order. He again found the recall of cued *S*s to be no greater than that of uncued controls. These findings were taken as evidence that item traces are not stored dependently.

Subsequent investigators have pointed out that effective cuing must meet certain minimal conditions. First, the list must be sufficiently organized in storage. This would presumably occur with minimal training on structured lists (e.g., categorized lists) or with extended training on unstructured lists (e.g., lists comprised of unrelated words). Second, the cues provided must be appropriate to the organization established, i.e., at least one member of each stored S unit should be provided. Third, the cues provided must supplement those *S* is able to provide for himself through recall; presumably, if *S* can recall at least one member from each S unit unaided, providing cues will be of no benefit. A number of investigators have recently demonstrated that cuing with list members facilitates recall when these three conditions are met (e.g., Allen, 1969; Luek, McLaughlin, & Cicala, 1971; Wood, 1969). In each of these demonstrations, highly structured lists were employed; parallel conditions, in which unstructured or weakly structured lists were employed (e.g., Allen, 1969; Wood, 1969), produced no such facilitation. This discrepancy leads to the rather unparsimonious implication that item traces are stored relationally when structured lists are employed, but independently when unstructured lists are employed. The purpose of the present experiment was to provide a more adequate test of the trace-dependent storage hypothesis in situations in which unrelated words are used.

A number of difficulties arise in meeting the criteria indicated earlier when unstructured lists are employed. One of these is the likely fact that *S*s form no more S units than they are able to retrieve unaided. The method of circumventing this difficulty in the present study was to interpolate training on a second list between extensive training on an original list and a test of retention for that list. Tulving and Psotka (1971) have demonstrated with categorized lists that interpolated learning (IL) produces retroactive inhibition (RI), which in their situation acted primarily on the accessibility of categories, without disrupting the integrity of the categories formed. That is, RI loss was in terms of the number of categories represented in recall and not in terms of the number of items representing each recalled category. Cuing with category labels eliminated this RI. If similar organizational processes are involved in the learning of lists of unrelated words, cuing should be effective in reducing RI in such situations as well.

Another difficulty inherent in the use of unstructured lists is the choice of list member cues appropriate to the organization formed. As indicated earlier, the cue words should include at least one member of each S unit formed. To accomplish this in the present experiment, cues were selected from odd-numbered serial positions on the final free-recall test of original learning (OL). This insures the presence of at least one member of each S unit among the cues provided, assuming that S-unit members are recalled consecutively. In addition, providing these cues in the order in which they occurred in recall should avoid any disruption of linkages that may have been established between S units.

---

[1] This research was supported in part by a faculty research grant awarded by the Fresno State College Foundation.

[2] Requests for reprints should be sent to David R. Basden, Psychology Department, California State University, Fresno, California 93710.

Method. The *S*s were 36 psychology students of both sexes serving in partial fulfillment of course requirements at California State University at Fresno.

The *S*s were tested individually in sound-isolated cubicles. The lists were presented in a Lafayette memory drum, and *S*s recorded their recall on lined paper provided by *E*. Two lists (A and B) of 30 common nouns were employed. Obvious intralist and interlist associations were avoided in preparing the materials. Seven random orders of each list were prepared.

Following standard free-recall instructions, the list was presented at a 2-sec. rate. The *S* read each word aloud as it appeared and following presentation of the last word engaged in 30 sec. of backward counting as a recency control. This in turn was followed by a 2-min. written free-recall test. This procedure was repeated with presentation order varied from trial to trial until *S* met a criterion of 21 of 30 correct responses plus one trial. Immediately following OL, *S* was told that he was to learn a second list, and the first trial of IL began. The same procedure was followed in IL as in OL with the exception that *S* was trained for seven trials. List A and List B were used equally often in OL and IL. During IL, *E* prepared a cued response sheet for a random half of the *S*s. This was accomplished by recording in appropriate order at the top of the recall sheet 10 words *S* had recalled on the final trial of OL. These were words recalled in odd-numbered serial positions 1–19. Immediately following the final recall test of IL, *S* counted backward for 30 sec. The *S* was then given either a blank recall sheet or the recall sheet containing cue words and was asked to recall as many words from the first list as he could remember. This retention test continued for 2 min.

Results. In OL, the mean numbers of trials given on List A for cued and uncued *S*s were 5.67 and 6.33, respectively, and for List B the means for cued and uncued *S*s were 5.67 and 6.56, respectively. An analysis of variance showed neither the main effect of list ($F < 1$) nor that of cuing condition, F (1, 32) = 2.17, $p > .10$, to be significant. The interaction between these two factors also failed to reach significance ($F < 1$). The overall mean number of correct responses on the final trial of OL was 23.80. An index of intertrial repetition (Bousfield & Bousfield, 1966), modified to include bidirectional repetitions, was calculated for the final two trials of OL for each *S*. The mean value of the index for cued *S*s was 3.19, indicating above chance organization, t (17) = 6.26, $p < .01$, and the mean value of the index for uncued *S*s was 3.03, again indicating above chance organization, t (17) = 6.46, $p < .01$. Cued and uncued *S*s did not differ in the level of organization achieved, t (34) = .22, $p > .10$. In IL the mean numbers of correct responses on Trial 7 for cued and uncued *S*s were 24.28 and 24.39, respectively, and this difference was not significant, t (34) = .02, $p > .10$. The 10 words which had occupied odd-numbered serial positions in the final recall test of OL were discounted for both groups of *S*s in the analysis of the retention test. These were the words used as cues for cued *S*s; the remaining list members were termed critical words. Mean critical word recalls for *S*s in the cued and uncued conditions were 11.56 and 8.94, respectively. This difference was significant, t (34) = 2.85, $p < .01$.[3] Only two words from the IL list appeared as intrusions in the retention test, with one appearing in the protocol of a cued *S* and the other in the protocol of an uncued *S*.

The overall magnitude of retention loss can be estimated from a comparison of critical word recall on the final trial of OL and the retention test. Mean loss for cued *S*s was 2.67, t (17) = 3.62, $p < .01$, and mean loss for uncued *S*s was 4.83, t (17) = 7.52, $p < .01$. The difference in mean loss was, of course, significant, t (34) = 2.66, $p < .05$.

Discussion. It is apparent that cued and control *S*s achieved equivalent performance during OL and maintained this equivalence in IL; hence, the superior recall of cued *S*s on the retention test can be attributed to the provision of list members as cues during that test. Although there was marked retention loss under both conditions, the magnitude of the loss was considerably less for cued *S*s than for control *S*s. An interpretation consistent with previous research is that IL reduced the accessibility of S units formed during OL, and that cuing reestablished access to those units (e.g., Tulving & Psotka, 1971). An alternative interpretation is that the associative bonds formed between the contextual stimulus and the S units during OL were unlearned during IL and that cuing provided the mediational stimuli for the recall of the members of the S units. The negligible intrusion rate makes any significant retention loss through either competition or breakdown in list differentiation unlikely.

It can be concluded that when the conditions are met for cuing to be effective, cued recall is higher with unstructured lists just as with structured lists. If it is accepted that the trace-dependent storage hypothesis and the trace-independent storage hypothesis lead to opposed predictions with regard to the effectiveness of cuing, the present results must be considered as evidence in favor of the dependent storage view; however, the validity of this divergence of predictions is open to question. There can be little disagreement regarding the existence of list organization, the controversy rather being centered on the temporal locus of that organization. The trace-dependent storage hypothesis is that organization takes place during list presentation and that items are stored in accordance with this organization. The trace-independence hypothesis, on the other hand, is that item traces are stored independently during list presentation and that organization is imposed during recall. In either case, the basis for list organization must derive from *S*'s past experience, i.e., must exist in the memory store in the form of direct or mediated associations among list members. Given the existence of a network of associations stored in memory and established prior to list presentation, the trace-dependent vs. trace-independent storage controversy may be restated as follows: Is the associational network utilized in the storage of a list of items or only in the retrieval of those items once stored? In either case, recalled list members or list members provided by *E* should have a facilitative effect on further recall. If traces are stored dependently, the cuing function is direct; however, if traces are stored independently, cues should facilitate recall by providing access to the associative network which mediates organization. In other words, cuing with list members should facilitate recall whether presented items are stored in an independent or a dependent manner. This is not to deny the value of those studies (including the present one) that have compared cued and uncued recall under various input and output conditions, for these have contributed heavily to our understanding of organizational processes in free-recall learning. It may be, however, that such studies cannot contribute information regarding the temporal locus of list organization.

---

[3] An alternative analysis for this type of design involves comparison of the proportion of eligible words (here, 20 for cued *S*s and 30 for uncued *S*s) recalled by the two groups. In the present study, the mean proportions were .58 and .50 for cued and uncued groups, respectively, $t(34) = 7.90$, $p < .001$, thus confirming the critical word analysis.

REFERENCES

ALLEN, M. Cueing and retrieval in free recall. *Journal of Experimental Psychology*, 1969, **81**, 29–35.

BOUSFIELD, A. K., & BOUSFIELD, W. A. Measurement of clustering and of sequential constancies in repeated free recall. *Psychological Reports*, 1966, **19**, 935–942.

LUEK, S. P., MCLAUGHLIN, J. P., & CICALA, G. A. Effects of blocking of input and blocking of retrieval cues on free recall learning. *Journal of Experimental Psychology*, 1971, **91**, 159–161.

SLAMECKA, N. J. An examination of trace storage in free recall. *Journal of Experimental Psychology*, 1968, **76**, 504–513.

SLAMECKA, N. J. Testing for associative storage in multitrial free recall. *Journal of Experimental Psychology*, 1969, **81**, 557–560.

TULVING, E., & PSOTKA, J. Retroactive inhibition in free recall: Inaccessibility of material available in the memory store. *Journal of Experimental Psychology*, 1971, **87**, 1–8.

WOOD, G. Retrieval cues and the accessibility of higher order memory units in multitrial free recall. *Journal of Verbal Learning and Verbal Behavior*, 1969, **8**, 782–789.

(Received May 25, 1972)

Journal of Experimental Psychology
1973, Vol. 98, No. 2, 432–434

DECISION LATENCIES TO THEMATIC AND NONTHEMATIC DISTRACTORS IN PROSE[1]

ALAN KENNEDY[2]

University of Dundee, Dundee, Scotland

Two groups of *S*s carried out a recognition task after reading short passages of highly metaphorical prose. For one group the passages were headed with a title which provided the possibility of organizing the metaphors into a continuous narrative stream. Latencies to reject distractors in a recognition task were recorded, with the hypothesis that possession of the theme should slow rejection of thematically related words. The hypothesis was confirmed, but only for the easier of two passages. The results are discussed in terms of the model of recognition memory proposed by Hall, Sekuler, and Cushman in 1969.

This article is concerned with the understanding and retention of short passages of highly metaphorical prose. The materials, originally developed by Dooling and Lachman (1971), can be comprehended as a whole only if *S* possesses a cue to the underlying theme; otherwise, they are apt to be seen as collections of understandable but only vaguely related sentences. Dooling and Lachman claimed that recall and recognition of those words relating to the theme of such passages were higher if *S* knew the theme prior to reading. They argue from this that the theme acts as a vital mnemonic device in the retention of prose in the sense of providing an organizing schema. Although this argument is plausible, there are a number of questions left unanswered. Their procedure involves a very low success rate and under these circumstances it is questionable whether recall is an appropriate index of the organizing effects of a theme. The results of their first experiment support this criticism: Recall was improved by presenting a theme title, but for *all* classes of words (i.e., regardless of whether they related to the theme), and the effect was as large when the words of the passage were presented randomly as when they were read in natural order. The second experiment examined recognition scores, showing a higher hit rate for thematically related words in groups given the theme. However, again recognition was improved as much by cuing with a title when the material was randomly presented as when read normally. The possible interpretation cannot be escaped that *S*s in this situation matched individual items *not*, as Dooling and Lachman argued, to the theme, but simply to the title itself. Error rates were easily high enough for the presence of a title to secure the obtained marginal improvement in hit rate.

The present study attempted to look more directly at the hypothesis proposed by Dooling and Lachman (1971). The materials were read, not word by word, but in as natural a fashion as possible, and the effects of precuing with a theme were assessed by looking at the false recognition rate and latencies of response to associates of the theme. This procedure, adapted from Hall, Sekuler, and Cushman (1969), has been shown to provide a sensitive index of the perceived "proximity" of associates to key items in lists and sentences (Kennedy, 1972). The procedure rests on a model of the recognition process which assumes that increments in the judged "familiarity" of words accrue as a result of their presentation in a particular context. In addition, though to a lesser extent, increments in the familiarity of associated items not presented also takes place. In a typical experimental situation *S* is assumed to establish a criterion on the "perceived familiarity" axis; if on any given trial perceived familiarity exceeds this value, the item is judged as old; otherwise it is seen as new. In terms of this model, the probability density functions for different classes of stimulus items lie at different points along the dimension of perceived familiarity. Response latency in a recognition task is a function of the distance between the perceived familiarity and the established criterion. If knowledge of the theme of a passage serves to produce a better organized associative structure on reading it, then this should be shown both in a higher false recognition rate to thematic associates and in retarded latencies of correct rejections.

Method. The *S*s were 20 students at the University of Dundee. They were unpaid volunteers, naive as to the purposes of the experiment.

The two short passages (entitled "Christopher Columbus Discovering America" and "The First Space Trip to the Moon"), devised by Dooling and Lachman (1971), were used without alteration. The following 32 nouns were selected from the two passages and used in a later recognition task: Passage 1—GEMS, HERO, LAUGHTER, SCHEME, EYES, EGG, TABLE, PLANET, SISTERS, PEAKS, DAYS, DOUBTERS, RUMORS, EDGE, CREATURES, and SUCCESS; Passage 2—QUARTERS, OBJECTS, BLACKNESS, FIELD, COMPANIONS, BUTTONS, PATTERNS, HOMELAND, BALL, THINGS, MOUNTAINS, VALLEYS, PAPERS, ARRIVAL, MAN, and NEWS. For each passage, an additional list of 16 words was constructed, consisting of 8 thematically related nouns and 8 unrelated nouns. The thematically related list for Passage 1 was:

---

[1] This research was supported by the Social Science Research Council. Grateful acknowledgment is made to Barbara Wilkinson who carried out the testing.

[2] Requests for reprints should be sent to R. A. Kennedy, Department of Psychology, University of Dundee, Dundee DD1 4HN, Scotland.

SHIPS, EARTH, BIRDS, SEA, STORM, WAVES, SAILORS, and VOYAGE; the list for Passage 2 was: ROCKET, LANDING, ORBIT, STARS, INSTRUMENTS, CRATER, METEOR, and SATELLITE. These words were chosen intuitively following a detailed analysis of each passage. The unrelated lists of words were chosen to be matched as far as possible for letter length and number of syllables. In the final lists used in the recognition task, each word appeared twice, the whole list being preceded by two fillers. Ten different randomized orders of the 64 words were used.

Two groups of 10 *S*s were tested. The procedure for both groups was identical except that *S*s in Group 1 were told the title of each passage before reading it and also received versions of the passages headed with a title. All *S*s read both passages aloud, order of presentation being balanced. After reading a passage the recognition task consisted of the 66 items relating to the passage presented at a 3-sec. rate, *S* being told to push a right-hand button for old words and a left-hand button for new words. Following this task the passage was read again and recognition was tested with another random order of the same 66 test items. This procedure was followed for the second passage with a new set of test items. In all, each of 10 different random orders of test items was used four times.

The passages were given to *S* typed in uppercase letters without punctuation on a sheet of paper, and his reading was tape recorded. The recognition test was controlled by a remote display interfaced to a PDP-12 computer. Items were displayed using a standard uppercase character display. Testing took place in an acoustic booth.

Results. The mean number of false decisions to old words for each *S* were subjected to an analysis of variance. Although the total number of errors was greater for Passage 1 than Passage 2 (360 vs. 313), this effect was not significant, F (1, 18) = 3.24.

The mean number of false positives (responding old to a new word) made in the two recognition tests on both passages are shown in Figure 1. Significant effects were shown for trial, F (1, 18) = 14.41, $p < .01$; type of item (thematically related or not), F (1, 18) = 23.45, $p < .01$; the Item Type × Cue interaction, F (1, 18) = 13.19, $p < .01$; the Trial × Item Type interaction, F (1, 18) = 8.79, $p < .05$; and the triple interaction, F (1, 18) = 11.70, $p < .05$.

Since the total number of false positives differed very greatly for the two passages, analyses were conducted on each separately. In Passage 1, all the effects above were significant; in Passage 2, a significant effect was shown for the Item Type × Cue interaction, F (1, 18) = 9.95, $p < .01$.

Mean response latencies to make correct old decisions were subjected to a three-way analysis of variance in which factors were presence of cue, passage, and trial: The only significant effect was for trial, latencies being faster on Trial 2, F (1, 18) = 66.21, $p < .01$.

The two passages were analyzed separately. Mean correct new decision latencies for the two passages are shown in Figure 2. For Passage 1 the only significant effects were for trial, F (1, 18) = 19.14, $p < .01$; and item type, F (1, 18) = 4.46, $p < .05$. The predicted Cue × Item Type interaction was not significant, F (1, 18) = .64. The analysis of Passage 2 showed significant effects for trial, F (1, 18) = 36.01, $p < .01$; item type, F (1, 18) = 6.08, $p < .05$; and the Cue × Item Type interaction, F (1, 18) = 6.51, $p < .05$.

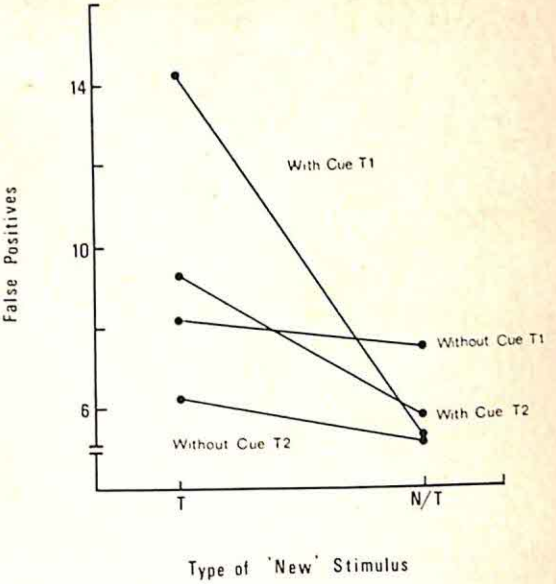


FIG. 1. Number of false positives to thematic (T) and nonthematic (N/T) distractors for both passages combined following the two readings (T_1 and T_2).

The total time to read each passage, excluding reading the title, was remarkably consistent for all subconditions of the design, lying between 33.5 and 34.3 sec. The tape recordings of these readings were analyzed for three types of errors: phrasing difficulties, relating to intonation patterns; repetition of words or parts of words; and misreadings. Although the number of cases was too small to allow further analysis, some interesting trends were apparent. Combining both passages, the total number of reading errors was greater for the with-cue group (63 vs. 47), but this difference was almost solely due to phrasing errors (36 vs. 14), the majority of which were contributed by data from Passage 1.

Discussion. Considering the error data first, no significant differences were found in the number of false categorizations of old stimuli in the various subconditions of the design. This finding contrasts with the results of Dooling and Lachman (1971, Experiment II), who showed large differences in scores (after a correction for guessing) on thematically important words in favor of the with-cue *S*s. Since the present study used as stimuli most of the nouns from each passage, and these are likely to have been judged as thematically significant, the results provide further support for the suggestion that some *S*s in Dooling and Lachman's situation achieved improvements in hit rate which could not be attributed to the organizing effects of the theme.

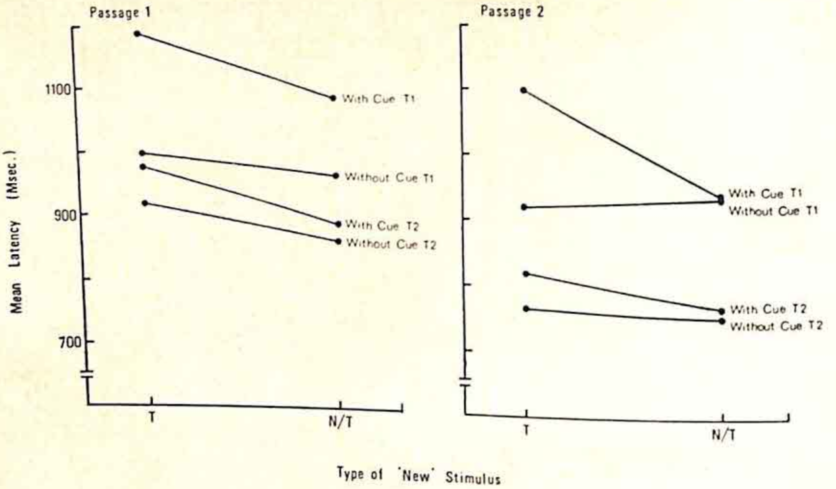


FIG. 2. Mean latency to reject thematic (T) and non thematic (N/T) distractors after each reading of the two pages.

The false recognition rate of words related to the theme shows much more clearly the influence of the cue. Although the total number of errors differed for the two passages, Passage 1 being more difficult, a Cue × Item Type interaction was present in both. Figure 1 shows quite clearly that thematically related items, particularly on the first trial, produce many more recognition errors than unrelated items.

In terms of the model outlined earlier it was predicted that an effective cue should so organize the passage that many associations would be recruited which would not be available to *S*s reading without a cue. The effect of this enlarged and more highly organized structure should be to delay the rejection of associatively proximal items. It should be stressed that this index is based on successful rejection and is not open to the criticism that the effects relate to guessing strategies. It is clear from Figure 2 and the related analyses that such a result obtains in the case of the second passage, where quite clearly the possession of a cue produces a relative delay in the rejection of thematically related stimuli, the effect being attenuated on the second trial.

In Passage 1 precuing with the theme appears simply to have generated a general slowing down of *all* responses on the first trial, producing as large an effect for nonthematic as for thematic stimuli. In view of the error data it seems unlikely that this result stems from inadequate differentiation between thematically related and unrelated items: False recognition rates were changed in the with-cue condition, but the latencies of successful rejections were unaffected. Also, for the same reason, failure to produce the expected Cue × Item Type interaction cannot have resulted simply from a poor choice of thematically related items. It is strongly suggested that the mechanisms producing changes in error rates and those underlying changes in rejection latency are not necessarily identical. The possession of a theme in the case of a difficult passage appears to increase errors to both types of test items, and particularly thematically related errors, but this may not be a reflection of some highly organized underlying schema. In fact, the data from reading errors suggest the reverse, that is, that the presence of theme in the case of a highly metaphorical passage produces a marked increase in phrasing errors. The *S* appears to take the cue as a challenge to organize the passage into appropriate syntactic units. In failing to do this he produces an increased number or errors and a poorer recognition score. The latency measure, based on successful rejection and providing a purer index of the degree of associative "priming" brought about by the theme, illustrates that *S*s may fail to understand a passage, even when cued by a theme. The latency data suggest that the effect shown by Dooling and Lachman (1971) may have been present only for the second, easier passage.

In summary, the methods used in the present study confirm that rejection latencies are a reliable index of certain strategies adopted in processing prose. The suggestion that possession of a cue leads to the establishment of an organizing surrogate process gains some support, particularly in the case of easy passages.

REFERENCES

DOOLING, D. J., & LACHMAN, R. Effects of comprehension on retention of prose. *Journal of Experimental Psychology*, 1971, **88**, 216–222.

HALL, J. W., SEKULER, R., & CUSHMAN, W. Effects of IAR occurrence during learning on response time during subsequent recognition. *Journal of Experimental Psychology*, 1969, **79**, 39–42.

KENNEDY, R. A. Semantic constraints on location operations in simple sentences. *British Journal of Psychology*, 1972, **63**, 387–393.

(Received May 25, 1972)

Journal of Experimental Psychology
1973, Vol. 98, No. 2, 435–437

EFFECTS OF IRRELEVANT INFORMATION IN SPEEDED DISCRIMINATION

HAROLD L. HAWKINS,[1] GARY J. McDONALD, AND ABIGAIL K. COX
University of South Florida

Test stimuli in a *same–different* reaction time (RT) task varied along two independent binary dimensions, size (S) and form (F). During Phase 1 (64 trials), both experimental and control groups were instructed to respond to S, ignoring F. During Phase 2 (196 trials), the members of the experimental group were told to respond on the basis of both S and F, whereas members of the control group were instructed to continue responding only to S. Phase 3 was identical to Phase 1 for both groups. Both *same* and *different* RTs were slower during Phase 1 when the irrelevant dimension was incompatible in state (*same* or *different*) to the relevant dimension. The magnitude of this effect was uninfluenced by the Phase 2 experience of the experimental group. These results are inconsistent with the view that irrelevant information has its effects in speeded discrimination through response conflict. Two perceptual interpretations were discussed.

Response latency in *same–different* reaction time (RT) is retarded when the relevant and irrelevant features of test stimuli are incompatible in state, i.e., when stimuli are *same* along relevant dimensions and *different* along irrelevant dimensions (Egeth, 1966) or vice versa (Hawkins & Shigley, 1972). Egeth suggested that this effect is mediated by response conflict. According to this view, when a particular feature serves as a relevant aspect of the stimuli to be discriminated, there develops a tendency to respond on the basis of the state of that feature and this tendency will persist beyond the time during which the feature is relevant. Thus if a stimulus pair contains two dimensions, one relevant and one not, if the relevant dimension on prior occasions has been relevant, and if the two dimensions are incompatible in state, then two opposing response tendencies will be evoked. Under these circumstances response accuracy requires the suppression of the irrelevant tendency. Presumably, it is the time needed to suppress this conflicting response tendency that constitutes the irrelevant information effect. The purpose of the present study was to test directly the foregoing interpretation.

Test stimuli contained two independent binary dimensions, denoted here as X and Y. The experimental group based its *same–different* judgments solely on the state of X during Phase 1, on both X *and* Y during Phase 2, and again solely on X during Phase 3. The control group based its judgments solely on X during all three phases. The response conflict notion implies that (*a*) there should be no effect of irrelevant information during Phase 1, since the irrelevant dimension had never previously been relevant, and (*b*) there should be an effect during Phase 3 for the experimental but not the control group, since the former uniquely experienced the now irrelevant dimension as relevant during Phase 2.

Method. The *S*s were 12 undergraduates at the University of South Florida who received course credit for participating.

Stimuli were viewed through a Polymetric two-field tachistoscope at a distance of 43.18 cm. On a desk in front of *S* were two response keys labeled *same* and *different*, respectively. Response latency was indicated by a Hunter Klockounter. Stimuli were gray trapezoids (Hawkins, 1969) constructed of cardboard and mounted on white index cards. Each trapezoid measured 16 mm. (height) × 15 mm. (short base) × 17 mm. (long base). The horizontally arrayed stimulus pair appearing on each trial subtended a visual angle of 6° on the horizontal and 2.4° on the vertical plane. Stimuli varied along two independent binary dimensions, size (S) and form (F). Dimension F was varied by orienting the long base either up or down. Each trapezoid contained an inked circle (No. 3 Rapidograph) either 7 or 9 mm. in diameter (Dimension S). Prior research (Hawkins & Shigley, 1972) has revealed that Dimension F, when irrelevant, has a reliable effect on *same–different* RT to Dimension S.

On Day 1, *S*s were given 60 practice trials on the *same–different* RT task with letter pairs as stimuli. Based on their performance on this task, *S*s were divided into blocks of slow and fast responders. Half of the *S*s from each block were then randomly assigned to the experimental condition and half to the control condition. On Day 2, all *S*s were given 64 *same–different* RT trials on which Dimension S was relevant and Dimension F was irrelevant (Phase 1). The stimulus deck used during this phase consisted of 32 cards: 8 on which stimuli were identical along S and F, 8 on which stimuli differed along F, 8 on which stimuli differed along S, and 8 on which stimuli differed along both dimensions. Thus, the *same* response was appropriate on one half of the trials. Experimental and control groups differed with respect to the Phase 2 task which was presented at the beginning of the experimental session on Day 3. The experimental group was re-

---

[1] Requests for reprints should be sent to Harold L. Hawkins, Department of Psychology, University of South Florida, Tampa, Florida 33620.

TABLE 1

MEAN *Same* AND *Different* REACTION TIMES AND PERCENTAGE OF INCORRECT RESPONSES (IN PARENTHESES)

Trial block	Experimental		Control	
	Compatible	Incompatible	Compatible	Incompatible
Same				
1[a]	671 (6.0)	709 (7.3)	676 (5.3)	716 (8.1)
2[b]	594 (4.9)	628 (4.8)	611 (4.5)	661 (7.5)
3	612 (5.5)	634 (7.0)	598 (2.8)	639 (7.5)
Different				
1[a]	705 (7.3)	748 (7.1)	697 (5.7)	755 (6.6)
2[b]	601 (5.0)	660 (6.5)	651 (5.5)	700 (6.4)
3	636 (4.6)	647 (5.6)	655 (3.1)	685 (5.6)

Note. In the experimental condition, both form and size were relevant during Phase 2; in the control condition, form was never relevant.

[a] Trial Block 1 was prior to Phase 2.

[b] Trial Block 2 was immediately following Phase 2.

quired during this phase to base their *same–different* responses on both S and F, i.e., stimuli had to be *same* on both dimensions for the *same* response to be correct. The stimulus deck used here was that used in Phase 1. Thus, the *same* response was appropriate on only one fourth of the trials—a device used to insure that *S* would attend to both S and F on all trials (Hawkins, 1969). The control group continued during Phase 2 responding only on the basis of F. However, the number of cards on which stimuli were *same* on F was appropriately reduced to equalize the probability of the *same* response (.25) across experimental and control conditions. Phase 2 consisted of 192 trials.

Phase 3 was identical to Phase 1; both groups received 64 trials where S was relevant and F was irrelevant. The *same* response was appropriate on 50% of the trials.

Throughout the experiment the intertrial interval (ITI) was 4 sec. At the conclusion of the ITI, a warning tone of 1-sec. duration was presented. Stimulus onset was coincident with the offset of this tone. The *S*s were informed of their latency following each correct response. Following each incorrect response, *S*s were told only that they had committed an error.

Results and discussion. Table 1 gives the mean, across *S*s, of the median latencies on correct *same* and *different* trials under the two experimental conditions. Three trial blocks were differentiated in the table. The first corresponds to Phase 1. The second and third correspond to the first and second half of the trials given during Phase 3. Phase 2 latencies, which are neither comparable across conditions nor pertinent to the focus of this study, were omitted.

The *same* and *different* data were each subjected to a separate randomized-block analysis of variance with premeasured *S* speed (fast or slow) and experimental condition as between-*S* variables, and trial blocks and irrelevancy condition (irrelevant dimension either compatible or incompatible in state with relevant dimension) as within-*S* variables. The analysis of *same* RT data revealed that latency decreased across trial blocks, $F\ (2, 16) = 12.58$, $p < .01$, was shorter for *S*s categorized as fast as opposed to slow, $F\ (1, 8) = 5.95$, $p < .05$, and was higher when the irrelevant dimension was incompatible, $F\ (1, 8) = 66.3$, $p < .01$. Fisher's least significant difference (LSD) test, with the criterion of significance set at .05, indicated that *same* RT was reliably higher under Trial Block 1 when the irrelevant dimension was *different* as opposed to *same* (LSD = 38 msec.). This finding, taken together with the overall significance of condition of irrelevancy and the failure to find an interaction between trial block and condition of irrelevancy, $F\ (2, 16) = .10$, $p > .10$, indicates that an irrelevant information effect existed within the *same* data at the outset of this experiment, and this effect was uninfluenced during Phase 3 by the Phase 2 experience of responding to Dimension F.

The same picture developed from the analysis of *different* RT data. Here, too, latency declined across trial blocks, $F\ (2, 16) = 14.6$, $p < .01$, was faster for *S*s categorized as fast, $F\ (1, 8) = 15.0$, $p < .01$, and was slower on trials where the irrelevant dimension was incompatible relative to compatible, $F\ (1, 8) = 26.8$, $p < .01$. The LSD test revealed that the difference between irrelevant–compatible and irrelevant–incompatible trials was reliable (LSD = 23 msec.).

The present results provided no support for the response conflict interpretation of the effects of irrelevant information in speeded discrimination. The results were consistent, however, with a memory-scan interpretation recently proposed by Hawkins and Shigley (1972) which assumes that (*a*) the physical comparison of simultaneously presented stimuli occurs at a relatively early stage in the information-processing sequence, prior to the time at which input makes contact with memory; (*b*) the determination of the relevancy of any particular feature of a complex stimulus requires a match between the representation of that feature of stimulus input and a stored list of the features defined as relevant; and (*c*) the increase in latency occurring when the states of relevant and irrelevant stimulus features are incompatible represents the time needed to scan memory to ascertain which dimension is relevant and, hence, which response alternative is appropriate. However, there is a perceptual counterpart to the response conflict interpretation that also adequately describes the present results. According to this view, the conflict arising out of state incompatibility operates to retard the speed with which stimuli can be perceptually categorized as either *same* or *different* rather than to retard the speed with which specific motor responses, e.g., key presses, learned during the experiment can be actuated. The association of perceptual sameness with the *same* category label and perceptual difference with the *different* label presumably is brought by *S* into the experimental situation. If this is the association with which irrelevant–incompatible information interferes, the interference effect should occur from the beginning of the experiment and be uninfluenced by the experience of the experimental group during Phase 2. The data of the present experiment do not distinguish between memory scan and category conflict interpretations. However, a weak comparison is possible based on data obtained in a recent experiment by Hawkins and Shigley (1972). The comparison hinges on the fact that speeded categorization time increases with the number of irrelevant stimulus dimensions, the number of relevant features being held constant (Hodge, 1959). The category conflict interpretation of this result is straightforward: The greater the number of stimulus features dictating responses incompatible with the correct one, the greater the category conflict and, hence, the slower the categorization time. The memory-scan interpretation is equally straightforward: Increases in the number of irrelevant stimulus features, provided at least one is incompatible with the relevant feature(s), lead to increases in the number of features whose relevancy must be established by means of a time-consuming scan of the stored list of those stimulus properties previously defined as relevant. With these interpretations in mind, consider a situation where the stimuli to be compared in *same–different* RT contain three dimensions, two relevant and one irrelevant. Of specific interest here are two cases: one where stimuli differ along one relevant dimension and along the irrelevant dimension, SD(D), and the other where stimuli differ only along the one relevant dimension, SD(S). According to the category conflict interpretation, *different* RT should be longer in the latter case because the total number of dimensions dictating the incorrect *same* response is greater, leading to greater conflict. The memory-scan interpretation predicts no difference between the conditions, because both exhibit state incompatibility and contain stimuli of equal complexity. A reanalysis of the Hawkins and Shigley (1972) data revealed that median latency on SD(D) trials was 502 msec. and on SD(S) trials was 497 msec. This difference, which in absolute value was opposite to that implied by the category conflict interpretation, did not approach significance, $t\ (5) = .37,\ p > .25$.

REFERENCES

Egeth, H. Parallel versus serial processes in multidimensional stimulus discrimination. *Perception & Psychophysics*, 1966, **1**, 256–262.

Hawkins, H. L. Parallel processing in complex visual discrimination. *Perception & Psychophysics*, 1969, **5**, 56–64.

Hawkins, H. L., & Shigley, R. H. Irrelevant information and processing mode in speeded discrimination. *Journal of Experimental Psychology*, 1972, **96**, 389–395.

Hodge, M. H. The influence of irrelevant information upon complex visual discriminations. *Journal of Experimental Psychology*, 1959, **57**, 1–5.

(Received June 19, 1972)

Journal of Experimental Psychology
1973, Vol. 98, No. 2, 438–439

STROOP INTERFERENCE WITH SUCCESSIVE PRESENTATIONS OF SEPARATE INCONGRUENT WORDS AND COLORS[1]

FREDERICK N. DYER[2] AND LAURENCE J. SEVERANCE

U.S. Army Medical Research Laboratory, Fort Knox, Kentucky

Color naming was delayed when a color patch was preceded by a 50-msec. flash of a color name in black ink. This delay was highly reliable, but only about half that obtained with individual stimulus versions of the Stroop color–word test where the word and color are integrally combined. Intervals of 0, 25, 50, and 100 msec. between the word and color did not change the amount of interference. The paradigm makes it possible to explore naming interference with visual dimensions that cannot be combined with words and, by manipulation of the interval between word and color, to study the duration of neural activity associated with words.

In the Stroop color–word test (Stroop, 1935) the time to name colors is greatly delayed for a condition in which the color patches are shaped to spell words denoting colors incongruent to the patch color. A similar delay of color naming has been found when incongruent color names in white are printed on small colored rectangles with the color of the rectangles to be named (Kamlet & Egeth, 1969). The present study was designed to determine if interference to color naming could be generated when the color and word are not integrally combined in either of these methods, but instead are presented sequentially, i.e., a black color name followed by a nonword color patch. If "Stroop interference" can be generated by such a procedure, this would open the door for exploration of the effects of irrelevant words on naming of other dimensions which, unlike color, cannot readily be combined with words. It would also allow study of the duration of the neural activity associated with a word stimulus. This could be done by manipulating the interval between presentations of the word and color and observing the effect on Stroop interference.

Method. The *S*s were eight males, ranging in age 18–24 yr., who volunteered for the experiment. Visual acuity, with correction if necessary, and color vision were normal.

The stimuli consisted of five word stimuli—the words *red*, *blue*, *green*, and *yellow* and a series of five Vs (alternately right side up and upside down) in black—and four color stimuli—a series of four Xs in each of the colors red, blue, green, and yellow. These nine stimulus cards were constructed with 48-point press-on letters (Prestype Futura Demi) centered on white cards. The black words were presented in Field 1 and the colored Xs were presented in Field 2 of a Scientific Prototype Model GB three-field tachistoscope. The *S* initiated the stimulus sequence by depressing a switch. The words were exposed for 50 msec., followed after either 0, 25, 50, or 100 msec. by the colored Xs which remained visible for 2 sec. A millisecond clock was started with exposure of the colored Xs and was stopped by the naming response through a voice key. A small fixation cross in the blank field remained on between stimuli including a return to this stimulus between the word and color for the 25, 50, and 100 msec. intervals between the black words and colored Xs. Background luminances of the words, colored Xs, and fixation cross were equated at 23 ftL.

All combinations of the five black-word stimuli and the four colored-X stimuli made a total of 20 different sequential stimulus pairs, with four control combinations (Vs followed by Xs), four congruent combinations of words and colors, and 12 incongruent combinations. At an experimental session, all 20 stimuli preceded by 4 randomly selected warm-up-trial stimuli were presented at one of the word-to-color intervals (0, 25, 50, or 100 msec.). Eight experimental sessions were given to each *S*, representing two repetitions of each word-to-color interval. The first and second four sessions each represented a repetition of the four intervals. A Latin-square constraint was imposed on the random order of intervals such that each word-to-color interval appeared twice in the eight successive sessions, when all eight *S*s' sessions were considered. The experimental sessions occurred on successive week days at approximately the same time each day for a particular *S*. The *S*s were instructed to name the color as quickly and as accurately as possible, and accuracy was further emphasized by providing feedback of the response latencies on only the correct trials. Stimuli on error trials were repeated later in the sequence to provide complete data. A 2-min. rest was provided midway in the experimental session. Time for the session averaged about 15 min. Prior to the experimental sessions a practice session of 24 stimuli with 6 stimuli at each of the four word-to-color intervals was given. The order of intervals and the order of the six stimuli at an interval were completely random and different for different *S*s at the practice session.

Results and discussion. Response latencies were analyzed in a repeated-measure analysis of variance with a four-level factor of color, a five-level factor of word, a four-level factor of interval, and a two-level factor of repetition. Color produced a highly signifi-

---

[1] The authors wish to thank Dennis M. Buck who collected the experimental data.

[2] Requests for reprints should be sent to Frederick N. Dyer, Experimental Psychology Division, U.S. Army Medical Research Laboratory, Fort Knox, Kentucky 40121.

cant main effect, $F\ (3, 21) = 6.35$, $p < .001$. Fastest color naming occurred for red with a mean naming time of 487 msec. Blue was named with an average time of 513 msec., yellow with 522 msec., and green with 531 msec. The time for naming red differed significantly (Newman-Keuls, $p < .05$) from the time for naming the other three colors, which did not differ significantly among themselves. A highly significant main effect also appeared for the word factor, $F\ (4, 28) = 9.48$, $p < .001$. Color naming for the condition with Vs as the stimulus was significantly faster (Newman-Keuls, $p < .01$) than for each of the other words. The other words did not differ among themselves. This illustrates faster naming for the control condition where color stimuli were preceded by the nonmeaningful Vs than for the condition where colors were preceded by words.

Of central interest is the Color × Word interaction which was highly significant, $F\ (12, 84) = 5.69$, $p < .001$. Mean naming time for the 12 different incongruent combinations was 533 msec., while for the congruent combinations the mean was 478 msec. These compare to the 488-msec. average naming time for control Vs. The average naming time for the incongruent combinations differed from the congruent and control conditions at a highly significant level (Newman-Keuls, $p < .01$), although the congruent and control conditions did not differ significantly between themselves. This 45-msec. difference in naming time between incongruent and control conditions is about half of the difference found between these conditions when words and colors were integrally combined (Dyer, 1971; Dyer & Kuehne, 1972).

One other significant interaction was Repetition × Word, $F\ (4, 28) = 3.97$, $p < .025$. This reflected a decrease in time to name colors when they were preceded by words from the first repetition to the second (528 msec. to 511 msec.) and a concurrent increase in the time to name colors when they were preceded by control Vs (484 msec. to 493 msec.). One possible explanation for this improvement in color-naming performance when the colors were preceded by words and the decline in performance on colors preceded by control Vs would be that Vs become relatively more novel and distracting after many stimulus presentations, four out of five of which are words. However, the naming time for Vs was not raised to that obtained with incongruent combinations.

No other main effects or interactions approached significance. This absence was surprising for the interval factor. In a previous study (Dyer, 1971) words were preexposed in black for various short intervals prior to coloration, and with as little as 100 msec. of black-word preexposure, interference to color naming was greatly reduced. In the present study it was expected that the 100-msec. interval between black-word offset and color onset would also produce a decline in interference. The fact that interference was almost constant for different word-to-color intervals suggests that the present 50-msec. word flash produced a longer lasting word representation than the continuous word presentations of the previous study.

The present paradigm, with longer intervals between words and colors, should permit study of the duration of central neural processes associated with word stimuli, including the effect on these processes of presenting various stimuli in the interval *between* the word and the color. In addition, modifications of this procedure, utilizing dimensions other than color, would allow comparisons of naming interference for a wide variety of dimensions, including those that cannot be integrally combined with words.

REFERENCES

Dyer, F. N. The duration of word meaning responses: Stroop interference for different preexposures of the word. *Psychonomic Science*, 1971, **25**, 229–231.

Dyer, F. N., & Kuehne, T. E. *Color naming latencies with brief exposures of individual Stroop and control stimuli.* (Tech. Rep. No. 970) Ft. Knox, Ky.: U. S. Army Medical Research Laboratory, 1972.

Kamlet, A. S., & Egeth, H. E. Note on construction of Stroop-type stimuli. *Perceptual and Motor Skills*, 1969, **29**, 914.

Stroop, J. R. Studies of interference in serial verbal reactions. *Journal of Experimental Psychology*, 1935, **18**, 643–662.

(Received June 22, 1972)

Journal of Experimental Psychology
1973, Vol. 98, No. 2, 440–442

SITUATIONAL GENERALITY OF CONDITIONING PHENOMENA: STIMULUS GENERALIZATION AND COMPONENT INTERACTION[1]

MICHAEL E. DAWSON[2] AND FRED W. DUNN[3]

Gateways Hospital, Los Angeles

Procedures modeled after those used in human Pavlovian conditioning experiments were employed in a probability learning paradigm. The conditioning phenomena of stimulus generalization and component interaction were shown to occur in the probability learning situation. These findings demonstrate the situational generality of conditioning phenomena and support the notion that human Pavlovian conditioning and probability learning may be mediated by similar processes.

The purpose of the present research was to determine whether certain Pavlovian conditioning phenomena occur in a simple probability learning situation. There were two related reasons for making this determination. The first reason was to test the situational generality of conditioning phenomena. As summarized at a conference of human learning specialists, "one of the most important classes of contributions needed . . . [is] . . . to test for some of the principal phenomena of conditioned-response learning . . . in other categories of human learning [Melton, 1964, p. x]." The second reason was to test the notion suggested by theorists of various persuasions (e.g., Estes, 1964; Humphreys, 1939; Tolman, 1959; Woodworth, 1958) that Pavlovian conditioning and probability learning may be mediated by fundamentally similar learning processes. If this notion is correct, then many of the same phenomena may be demonstrable with both paradigms.

The first experiment tested whether the conditioning phenomenon of stimulus generalization occurs in a probability learning situation. The procedures used to test for stimulus generalization were modeled after those employed by Hovland (1937) to study generalization of Pavlovian conditioned galvanic skin responses (GSRs). The second experiment tested whether the conditioning phenomenon of component interaction occurs in a probability learning situation. The procedures used to test for component interaction were similar to those employed by Grings and O'Donnell (1956) with Pavlovian conditioned GSRs.

Experiment I. The *S*s were 24 paid volunteers recruited from employment offices on college campuses. They were individually seated approximately 1 m. in front of two 15 × 15 cm. translucent screens. The left screen was nonoperative during the first experiment and will not be referred to again until the second experiment. Below the screens was a loudspeaker through which tones could be presented. Hunter timers, a Gerbrands programmer, and manual switches controlled the following sequence and duration of events: 500 msec. of tone, a 700-msec. "silent" interval, 1,000 msec. of screen illumination, and a 9,000-msec. intertrial interval.

The *S*s were instructed to predict, by means of an indicator dial, the occurrence or nonoccurrence of screen illumination immediately upon hearing a tone. The indicator dial consisted of a continuous linear potentiometer mounted inside a small metal casing. The potentiometer settings were controlled by a knob mounted on the metal casing. The middle setting of the dial was labeled "start," one extreme end setting was labeled "certain of light," and the other extreme end setting was labeled "certain of no light." The *S*s were instructed to set the knob at start during each intertrial interval and then, immediately upon hearing a tone, to turn the knob to the position which best described their prediction of screen illumination.

Following the instructions, the experiment consisted of three phases: adaptation, acquisition, and test for stimulus generalization. Adaptation consisted of the presentation of two blocks of four different tones each. The four tone frequencies of 200, 500, 1,000 and 2,000 Hz. were chosen so as to approximate those used in previous Pavlovian conditioning studies of stimulus generalization (e.g., Hovland, 1937). The orders of the tone presentations during adaptation were arranged by the method of incomplete counterbalancing (D'Amato, 1970, p. 55). The *S*s were told that this was merely an apparatus check and that there would be no screen illuminations during this phase. Acquisition then consisted of 30 presentations of a single tone frequency with 80% of the presentations followed by screen illumination (nonillumination occurred on Trials 7, 8, 13, 19, 23, and 28). The frequency of the tone presented during acquisition was 200 Hz. for half of the *S*s and 2,000 Hz. for the other half. The test for generalization then immediately followed acquisition and consisted of the presentation of two blocks of the four tones each. The orders of tone presentation during the generalization test were the same as those used during adaptation. No screen illumination was administered during the test for generalization.

---

[1] This research was supported by Public Health Service Grants MH-18411 and MH-17039. The authors wish to thank William W. Grings for use of equipment, Paul Reardon for technical assistance, and Michael Biferno, William W. Grings, James H. Satterfield, and Anne Schell for helpful comments on earlier versions of this paper.

[2] Also at California State University at Los Angeles.

Requests for reprints should be sent to Michael E. Dawson, Andrew Norman Research Center, Gateways Hospital, 1891 Effie Street, Los Angeles, California 90028.

[3] Now at Juvenile Rehabilitation Services, Lake View Terrace, California.

Since response magnitude is the typical dependent variable in Pavlovian autonomic conditioning experiments, while response probability is the usual dependent variable in probability learning experiments, a response measurement technique was devised which would yield both types of data. The technique consisted of having the potentiometer of the indicator dial control pen deflections of a Grass Model 7 polygraph. The range of pen deflections was from -23 mm. (certain of no light) to $+23$ mm. (certain of light). Thus, each response had a directional component and a magnitudinal component. For the response probability measure, each response was scored as either a prediction of "illumination" or "nonillumination," based solely on the directional component. For the response magnitude measure, a constant of $+23$ was added to each response, and the adjusted values, with a range 0–46 mm., were analyzed.

For purposes of data analysis, the three auditory test stimuli (S_1, S_2, and S_3) were arranged in order of their similarity to the original training stimulus (S_0). Table 1 shows the response probability and response magnitude data averaged across both blocks of test trials. Significant stimulus generalization was exhibited with both response measures. For the response probability data, a series of sign tests revealed that predictions of illumination occurred more frequently to S_0 than to any of the test stimuli and that they occurred more frequently to S_1 than S_3, $x\ (10) = 1$, $p = .011$. As for the response magnitude data, a series of Wilcoxon tests demonstrated that the mean response to S_0 was significantly larger than those to each of the test stimuli and that the response to S_1 was significantly larger than that to S_3, $T\ (23) = 56.5$, $p < .01$.

Experiment II. The *S*s were 40 paid volunteers recruited from employment offices of college campuses. The apparatus, procedures, stimulus durations, and instructions were similar to those used in the first experiment. However, instead of tones, colored lights were used as the predictive cues. The colors of red, blue, green, and yellow were formed by placing gels (Rosoline No. 6807, 6821, 6851, and 6871) in four equal sized sections of the left screen. The *S*s were instructed to predict the occurrence or nonoccurrence of right-screen illumination immediately upon seeing a light appear in the left screen.

Following the instructions, the experiment consisted of three phases: adaptation, acquisition, and test for component interaction. Adaptation consisted of the presentation of two blocks of the four colored lights each. The *S*s were instructed that this was merely to acquaint them with the lights and that no right-screen illumination would be presented during this phase. Acquisition consisted of 40 trials, 10 of each of the colored lights. Two of the colored lights (S +) were always followed by right-screen illumination, and the remaining two colored lights (S −) were never followed by illumination. The presentation order during acquisition was arranged randomly in blocks of four, and each color was assigned equally often to each reinforcement condition. The test for component interaction then immediately followed acquisition and consisted of the presentation of two blocks of the following five stimuli: S +, S −, S + S +, S + S −, and S − S −. The presentation order during the component interaction test was determined by incomplete counterbalancing, and no right-screen illumination was presented during this phase.

TABLE 1

Probability and Magnitude Data Obtained during Test for Stimulus Generalization

Response measure	Stimulus			
	S_0	S_1	S_2	S_3
Probability	.79	.46	.33	.25
Magnitude	36.6	20.6	15.2	12.9

TABLE 2

Probability and Magnitude Data Obtained during Test for Component Interaction

Response measure	Stimulus		
	S + S +	S + S −	S − S −
Probability	.58	.44	.21
Magnitude	27.2	19.3	9.7

Based on component interaction data obtained in previous Pavlovian conditioning experiments, it was predicted that the responses to the compound test stimuli would be ordered: S + S + > S + S − > S − S −. The response probability and response magnitude data averaged across both blocks of test trials are shown in Table 2. Significant component interaction was exhibited with both response measures. As for the response probability data, a series of sign tests demonstrated that illumination was predicted more frequently to S + S + than S + S −, $x\ (19) = 5$, $p = .03$, and more frequently to S + S − than S − S −, $x\ (29) = 6$, $p = .007$. As for the response magnitude data, a series of Wilcoxon tests demonstrated that the mean response to S + S + was significantly larger than that to S + S −, $T\ (32) = 138.5$, $z = 2.35$, $p < .01$, and that the response to S + S − was significantly larger than that to S − S −, $T\ (33) = 110.5$, $z = 3.04$, $p < .01$.

Discussion and conclusion. The data reported here demonstrate that at least two Pavlovian conditioning phenomena, stimulus generalization and component interaction, occur in a probability learning situation. These data indicate that conditioning phenomena have a generality of application beyond the confines of the traditional conditioning paradigm. For example, the data suggest that conditioning principles may be generalized to situations which lack a clearly definable UCS and UCR. Also, the results are consistent with the theoretical viewpoint that Pavlovian conditioning and probability learning (and perhaps other paradigms which include repetitive stimulus pairings) may share common learning processes. If true, this means that the processes may be investigated with either paradigm and that a single theoretical system might be applicable to both types of learning. This is not to imply that the underlying processes which occur in the two paradigms are identical, but it does suggest that there may be much to be gained by more closely examining their commonalities.

REFERENCES

D'Amato, M. R. *Experimental psychology: Methodology, psychophysics and learning.* New York: McGraw-Hill, 1970.

Estes, W. K. Probability learning. In A. W. Melton (Ed.). *Categories of human learning.* New York: Academic Press, 1964.

Grings, W. W., & O'Donnell, D. E. Magnitude of response to compounds of discriminated stimuli. *Journal of Experimental Psychology,* 1956, **52**, 354–359.

Hovland, C. I. The generalization of conditioned responses: I. The sensory generalization of conditioned responses with varying frequencies of tone. *Journal of General Psychology,* 1937, **17**, 125–148.

Humphreys, L. G. Acquisition and extinction of verbal expectancies in a situation analogous to conditioning. *Journal of Experimental Psychology,* 1939, **25**, 294–301.

Melton, A. W. Preface. In A. W. Melton (Ed), *Categories of human learning.* New York: Academic Press, 1964.

Tolman, E. C. Principles of purposive behavior. In S. Koch (Ed.), *Psychology: A study of a science.* Vol. 2. New York: McGraw-Hill, 1959.

Woodworth, R. S. *Dynamics of behavior.* New York: Holt, 1958.

(Received August 14, 1972)

Manuscripts Accepted for Publication in the *Journal of Experimental Psychology*

Cross-Modality Transfer of Differential GSR Conditioning to Word Stimuli. Irwin J. Mandel (Albert Einstein College of Medicine, Yeshiva University, 1300 Morris Park Avenue, Bronx, New York 10461) and Wagner H. Bridger.

Stimulus Meaningfulness, Encoding Variability, and the Spacing Effect. Ovid J. L. Tzeng (201 Social Science B Building, Pennsylvania State University, University Park, Pennsylvania 16802).

Effects of Experimental and Preexperimental Organization on Recognition: Evidence for Two Storage Systems in Long-Term Memory. D. J. Herrmann (Department of Psychology, Stanford, University, Stanford, California 94305) and John P. McLaughlin.

Relation between Disjunctive Reaction Time and Stimulus Difference. Dwight W. Curtis (Department of Psychology, California State University, Fullerton, California 92634), Manley A. Paulos, and Stanley J. Rule.

Uncertainty Effects on Time to Access the Internal Lexicon. Roy Lachman (Department of Psychology, University of Kansas, Lawrence, Kansas 66044).

Encoding and Retrieval from Long-Term Storage. Gordon Wood (Department of Psychology, Olds Hall, Michigan State University, East Lansing, Michigan 48823) and Joyce Pennington.

Recognition Effects of Study Organization and Test Context. Larry L. Jacoby (Department of Psychology, Old Botany Hall, Iowa State University, Ames, Iowa 50010) and Reginald L. Hendricks.

Clarity versus Identificability of Repeatedly Flashed Patterns. Daniel J. Weintraub (Department of Psychology, Human Performance Center, University of Michigan, 330 Packard Road, Ann Arbor, Michigan 48104) and John A. McNulty.

Learning Relative Frequency Distributions: Some Perceptual and Cognitive Factors. Charles A. J. Vlek (Department of Psychology, University of Leiden, Stationsplein 10 (2de), Leiden, The Netherlands) and Hans H. Werner.

Retroactive Inhibition in Two Paradigms of Negative Transfer. Isabel M. Birnbaum (School of Social Sciences, University of California, Irvine, California 92664).

The "Twenty-Questions" Technique: Prediction of Visual Threshold and Measurement of Redundancy. Robert H. Keen (Department of Psychology, New York University, 4 Washington Place, Room 976, New York, New York 10003).

Organization by Modality in Short-Term Memory. Lars-Göran Nilsson (Department of Psychology, University of Uppsala, Svartbäcksgatan 10, S-753 20 Uppsala, Sweden).

Handedness and Adaptation to Visual Distortions of Size and Distance. S. M. Luria (Code 432, Submarine Medical Research Laboratory, Naval Submarine Base New London, Groton, Connecticut 06340) Christine L. McKay, and Steven H. Ferris.

Recognition and Cued Recall of Idioms and Phrases. Leonard M. Horowitz (Department of Psychology, Stanford University, Stanford, California 94305) and Leon Manelis.

Response Force as an Indicant of Conflict in Double Stimulation. Barry H. Kantowitz (Department of Psychological Sciences, Purdue University, Lafayette, Indiana 47907).

(*Continued on page 443*)

(*Continued from page 442*)

Effects of Taxonomic Instances as Implicit Associative Responses on Verbal Discrimination Learning. Frank E. Fulkerson (Department of Psychology, Western Illinois University, Macomb, Illinois 61455) and Lawrence A. Prindaville.

Delayed Recognition Testing, Incidental Learning, and PI Release. Stephen T. Carey (Department of Psychology, University of Alberta, Edmonton, Alberta, Canada).

Anagram Solving as a Function of Word Imagery. Kathleen Dewing (Department of Psychology, University of Western Australia, Nedlands, Western Australia 6009, Australia) and Paul Hetherington.

Verbal Discrimination Learning of Items Read in Text Material. Eugene B. Zechmeister (Psychology Department, Loyola University, 6525 North Sheridan Road, Chicago, Illinois 60626), Jack McKillip, and Stan Pasko.

Evidence that Recall-Anticipation Differences Increase with Increasing Time Intervals. William F. Battig (Institute for the Study of Intellectual Behavior, University of Colorado, Boulder, Colorado 80302).

When is Recall Spectacularly Higher than Recognition? Michael J. Watkins (Department of Psychology, 333 Cedar Street, Yale University, New Haven, Connecticut 06510).

Reminding and Retrieval from Permanent Storage. Herman Buschke (Department of Neurology, Albert Einstein College of Medicine, 1300 Morris Park Avenue, Bronx, New York 10461), Philip C. Goldberg, and Gerald Lazar.

Evidence for an Attentional Explanation of the Overtraining Reversal Effect. Frank D. Turrisi (Department of Psychology, Florida State University, Tallahassee, Florida 32306).

Effects of Word Frequency and Acoustic Similarity on Free-Recall and Paired-Associate Recognition Learning. Stephen W. Holborn (Department of Psychology, Boston University, 64 Cumming Street, Boston, Massachusetts 02215), Karen L. Gross, and Pamela A. Catlin.

Rate of Information Processing and Skin Resistance. Gordon B. Harding (Department of Psychology, New Mexico State University, Box 3452, Las Cruces, New Mexico 88001), Elizabeth E. Stevens, and Paul T. Marston.

Effects of Differences in Stimulus Retention on Concept Identification. William H. McVaugh (Department of Psychology, Old Dominion University, Norfolk, Virginia 23508).

Imagery and Theme in the Recall of Connected Discourse. Peter A. de Villiers (Department of Psychology, William James Hall, Harvard University, 33 Kirkland Street, Cambridge, Massachusetts 02138).

Perception of Letters, Words, and Nonwords. Dominic W. Massaro (Department of Psychology, Psychology Building, University of Wisconsin, Charter at West Johnson, Madison, Wisconsin 53706).

Implicit Speech in Reading Reconsidered. Stuart T. Klapp (Department of Psychology, California State University, Hayward, California 94542), Wallace G. Anderson, and Raymond W. Berrian.

Massed and Distributed Item Repetition in Verbal Discrimination Learning. Donald S. Ciccone (Department of Psychology, Rice University, Houston, Texas 77001).

Retention of Single-Cue Probability Learning Tasks as a Function of Cue Validity, Retention Interval, and Degree of Learning. Berndt Brehmer (Department of Psychology, Umea University, Umea 6, Sweden) and Lars-Åke Lindberg.

On the Locus of the Relative Frequency Effect in Choice Reaction Time. Harold L. Hawkins (Department of Psychology, University of South Florida, Tampa, Florida 33620), Stephen L. MacKay, Susan L. Holley, Bruce D. Friedin, and Stephen L. Cohen.

Special Announcement

Each year many of the manuscripts submitted to the *Journal of Experimental Psychology* are reviewed by referees whose names are not listed on the Masthead of the *Journal* as regular Consulting Editors. In 1972, 98 manuscripts were thus reviewed. The names of 6 of these reviewers now appear on the 1973 Masthead and 14 more will be added to the Masthead in 1974. In the meantime, the Editors of the *Journal* would like to express their appreciation of the services of these reviewers, 76 in number, who helped us and the authors by their competent and thorough reviews. Critiques of such quality are one of the most valuable means by which we as authors can improve our scientific communication and, in many instances, improve our research methods. And so the Editors of the *Journal* wish to thank the consultants whose names are listed below.

Earl A. Alluisi
Moshe Anisfeld*
Malcolm D. Arnoult
Harry P. Bahrick
William P. Banks
Ina McD. Bilodeau
Isabel Birnbaum
Michael H. Birnbaum
Robert A. Bjork*
Alexander M. Buchwald
Joseph C. Campione
Gerald L. Clore
Dwight W. Curtis
James H. Davis
Henry Dee
William N. Dember
Donald J. Dickerson
Michael E. Doherty
John W. Donohoe
Rachel Falmagne
Robert Fox
Peter A. Frey
Wendell R. Garner
Clifford B. Gillman
David M. Green
Anthony G. Greenwald
G. Robert Grice
Sherman L. Guth
David L. Hamilton
Douglas Hintzman*
Betty J. House
Janellen Huttenlocher
Leonard E. Jarrard
Mari R. Jones
Barry H. Kantowitz
William E. Kappauf*
Steven W. Keele
Deborah G. Kemler
Neal E. A. Kroll
Joseph S. Lappin*
Alvin M. Liberman
Gregory R. Lockhead*
Steven F. Maier*
Dominic W. Massaro*
P. D. McCormick
Richard B. Millward*
Jerome L. Myers*
Thomas M. Ostrom
Allen Parducci
Herbert L. Pick, Jr.
John R. Platt
Frank Restle*
Jack Richardson
Herbert Rubenstein
Peter H. Schiller
Marian Schwartz*
Keith G. Scott
Robert W. Sekuler
James C. Shanteau
Harvey G. Shulman*
Arnold M. Small, Jr.
Edward E. Smith
Kirk H. Smith
Terry J. Spencer*
George E. Stelmach
Manard R. Stewart
Bonnie Z. Strand
James M. Swanson
Robert Teghtsoonian
John Theios
Willard R. Thurlow
Constantine Trahiotis*
Don A. Trumbo
Michael T. Turvey*
Roy Wittlinger
Keith A. Wollin

---

* Reviewed more than one manuscript.

CAUTION:

Image 236 is NO longer
present in scan-order!

Maybe it was deleted from
outside of BCS-2 ?

CAUTION:

Image 237 is NO longer present in scan-order!

Maybe it was deleted from outside of BCS-2 ?

CAUTION:

Image 238 is NO longer present in scan-order!

Maybe it was deleted from outside of BCS-2 ?

CAUTION:

Image 239 is NO longer present in scan-order!

Maybe it was deleted from outside of BCS-2 ?

CAUTION:

Image 240 is NO longer
present in scan-order!

Maybe it was deleted from
outside of BCS-2 ?

CAUTION:

Image 241 is NO longer present in scan-order!

Maybe it was deleted from outside of BCS-2 ?

CAUTION:

Image 242 is NO longer present in scan-order!

Maybe it was deleted from outside of BCS-2 ?

CAUTION:

Image 243 is NO longer
present in scan-order!

Maybe it was deleted from
outside of BCS-2 ?

CAUTION:

Image 244 is NO longer present in scan-order!

Maybe it was deleted from outside of BCS-2 ?

CAUTION:

Image 245 is NO longer present in scan-order!

Maybe it was deleted from outside of BCS-2 ?

CAUTION:

Image 246 is NO longer present in scan-order!

Maybe it was deleted from outside of BCS-2 ?

CAUTION:

Image 247 is NO longer present in scan-order!

Maybe it was deleted from outside of BCS-2 ?

CAUTION:

Image 248 is NO longer present in scan-order!

Maybe it was deleted from outside of BCS-2 ?

CAUTION:

Image 249 is NO longer present in scan-order!

Maybe it was deleted from outside of BCS-2 ?

CAUTION:

Image 250 is NO longer present in scan-order!

Maybe it was deleted from outside of BCS-2 ?